DEVELOPMENT AND APPLICATION OF SYNTHETIC BIOLOGY TOOLS FOR NATURAL PRODUCT RESEARCH

BY

HENGQIAN REN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Doctoral Committee:

Professor Huimin Zhao, Chair
Associate Professor Mary L. Kraft
Professor Brendan A. Harley
Professor Wilfred A. van der Donk

Abstract

Natural products, or secondary metabolites, usually refer to bioactive small molecules produced by microorganisms or plants. To date, thousands of natural products have been characterized which reveal diverse structures and bioactivities, making it an ideal resource for discovering drug leads and mining biocatalysts. The advances in bioinformatics and DNA sequencing technology in the 21st century enables the prediction of tremendously large numbers of biosynthetic pathways for natural products directly from genomic database, which indicates Nature's great potential of making these molecules remains to be explored. However, most of these uncharacterized pathways are naturally under negative regulation or encoded by uncultivable microorganisms and thus turn into "dark matters" for natural product discovery.

Pathway refactoring serves as an invaluable synthetic biology tool for natural product discovery, characterization, and engineering. However, the complicated and laborious molecular biology techniques largely hinder its application in natural product research, especially in a high-throughput manner. Therefore, I developed a plug-and-play pathway refactoring workflow for high-throughput, flexible pathway construction, and expression in both *Escherichia coli* and *Saccharomyces cerevisiae*. The modular design not only enables the system to accommodate pathways with different number of genes, but also facilitates gene deletion and replacement. As proof of concept, a total of 96 pathways for combinatorial carotenoid biosynthesis were built successfully.

The plug-and-play workflow was then used for discovering lanthipeptides and glycocins, both are groups of ribosomally synthesized post-translationally modified peptides (RiPPs) with various

antimicrobial activities. A Class IV lanthipeptide with an unusual ring topology was successfully discovered, and its cyclization mechanism was elucidated *in vitro*. In addition, four novel glycocins were also discovered and one of them was unprecedentedly di-glucosylated on a single serine. Further bioactivity characterization of glycocins revealed that three of them exhibit narrow antimicrobial spectrum and implied the existence of new biological targets of glycocins.

In addition, I also characterized the biosynthesis of heat stable antifungal factor (HSAF), a polyketide which belongs to the polycyclic tetramate macrolactam (PTM) series. Previously Zhao group discovered a PTM from *Streptomyces griseus* and characterized its biosynthesis, indicating a parallel biosynthetic mechanism. To further explore if this is general for the biosynthesis of all PTMs, I made four constructs which are supposed to make different intermediates in the HSAF biosynthesis. After successful isolation and structural characterization of these intermediates, the biosynthetic mechanism of HSAF was elucidated, which clearly supports the assumption that the HSAF pathway also has a parallel biosynthetic mechanism.

Finally, to accelerate the engineering of streptomycetes to be ideal chassis for natural product discovery and overproduction, I applied the clustered regularly interspaced short palindromic repeats (CRIPSR)-dCas9 system for gene repression in both *Streptomyces lividans* 66 and *Streptomyces albus* J1074. By fusing the RNA polymerase omega factor (RNAP-ω) to dCas9, it was able to activate the silent undecylprodigiosin (RED) gene cluster in *S. lividans* 66. In addition, to further engineer the CRISPR-dCas9 system to exert titratable repression, an inducible promoter was used for small guide RNA (sgRNA) expression and its performance was analyzed.

Acknowledgements

First of all, I would like to thank my advisor Prof. Huimin Zhao for offering me the great opportunity to spend my past six years in his group. His continuous guidance, support and encouragement throughout my graduate studies helped me overcome all the challenges in my research and find the beauty of science. I would also like to thank Prof. Wilfred van der Donk for his support and collaboration in my RiPP discovery and characterization projects. In addition, I would express my appreciation to my doctorate committee members Prof. Mary Kraft, Prof. Brendan Harley and Prof. Wilfred van der Donk for providing me insightful comments to my research work.

Meanwhile, I would like to thank Dr. Yunzi Luo and Dr. Carl Denard for training me in my early days in the lab. I want to thank my collaborator Subhanip Biswas for his contribution to my glycocin project. I also want to thank my undergraduate student Sherri Ho who has worked with me on several projects for almost three years. I'm also grateful to other former and current members in Zhao group including Dr. Tong Si, Dr. Ryan Cobb, Dr. Todd Freestone, Dr. Jing Liang, Dr. Jiazhang Lian, Ran Chao, Dr. Lu Zhang, Dr. Zehua Bao, Dr. Yajie Wang, Dr. Bin Wang and Dr. Mingfeng Cao for their insightful discussion on my research and help in my daily life. I would also like to thank Lucas Li, Peter Yau and Brian Imai from Roy J. Carver Biotechnology Center for technical support.

Finally I would like to thank my parents for their unconditional love and supports in my life. I would also like to thank all my friends for their companies and all the happiness they bring to me throughout my life in Champaign.

Table of Contents

Chapter 1: Overview for Natural Product Discovery Through Synthetic Biology Approaches

1.1 Natural Product Discovery Through Genome Mining

Since the discovery of penicillin by Fleming in 1928, natural products have been playing an indispensable role in the treatment of human, animal and plant diseases for almost a century (1, 2). Natural products, particularly secondary metabolites from microorganisms, offer their producers advantages for growth and propagation in complex environments. For example, antibiotics, which can kill or stop the growth of a broad or narrow spectrum of target organisms, enable their producers to be more competitive for nutrients. This unique feature of antibiotics was soon explored in clinic to control pathogens. Today, with thousands of natural products being discovered with various bioactivities, natural products and their semisynthetic derivatives are widely used for the treatment of multiple kinds of diseases ranging from cancer to high cholesterol. Besides the treatment of human diseases, natural products are also used in the agriculture sector as herbicides, insecticides, and fungicides (3).

Traditionally, natural products were discovered through a combination of separation techniques and bioactivity screening. Using this strategy, the discovery of natural products entered its Golden Age in the period from 1950s to 1960s. However, since then, the rate for natural products discovery has been significantly decreased due to the inherent drawbacks of the traditional method. One of its biggest drawbacks is the high rediscovery rate (4). Importantly, it was found that there are numerous silent natural product biosynthetic gene clusters (BGCs) in a wide variety of microorganisms. The biosynthesis of natural products is usually under complex yet unknown regulations and therefore cannot happen under inappropriate culturing conditions. With the recent

advances of next generation sequencing and bioinformatic tools for BGC prediction, it has become clear that the potential of Nature to produce natural products was largely underestimated in the past. Take Streptomyces species as an example, around 20-50 putative BGCs can be predicted from each genome, while only a few of them have been characterized (5, 6). Moreover, as natural products potentially produced by unculturable microorganisms, such as some symbionts of marine animals and plants, are largely hindered from being discovered due to the limited availability of materials, the number of natural products that can be produced by uncharacterized BGCs are far more than the number of known natural products.

To activate silent BGCs as well as to make the natural product discovery process more efficient, many synthetic biology tools have been developed in the past decade. In this chapter, I will introdcue the synthetic biology strategies and tools developed for activating silent BGCs for discovery of natural products (Figure 1.1). In addition, I will also provide some background about the classification of natural products and mainly focus on introducing ribosomally synthesized and post-translationally modified peptides (RiPPs) and polyketides. Tools and strategies for optimizing the productivity of natural products will also be mentioned briefly.

1.2 Synthetic Biology Approaches for Natural Product Discovery

1.2.1 Advantages for Natural Product Discovery Through Heterologous Expression of BGCs

There are generally two strategies to address the limitations of traditional genome mining methods mentioned above. One is to turn on the silent BGCs through genetic manipulation of the native producer to rewire the corresponding negative regulation. While the other one is to heterologously

express the BGC in a model organism through direct cloning or refactoring (7). Although many natural products were successfully discovered through manipulation of native producers, there are several shortcomings associated with this strategy. For example, for BGCs in microorganisms lacking proper culturing methods in the laboratory, the isolation of target compounds through cell cultivation can hardly be achieved. Developing a cultivation method for a specific microorganism is also challenging because the nutrients critical for its survival are not predictable and mimicking the environment from where the microorganism was isolated seems to be a daunting task. Moreover, although many transformation or conjugation methods and broad host range genetic elements are available, the lack of genetic tools for most microorganisms remains the biggest challenge for activating silent BGCs in native producers.

However, such issues can be readily solved by expressing the BGC of interest in a heterologous host which provides a similar *in vivo* environment for successful gene expression and product synthesis as the native producer yet has well-established genetic manipulation tools for cloning, refactoring, and introduction of foreign BGCs. These good features not only make heterologous expression associated methods as alternative ways to native producer associated methods, but also shed light on developing general workflow for genome mining as all experiments involved have the potential to be standardized and therefore automatable. In light of the rapidly developing automation technologies for synthetic biology as well as the automation-based high-throughput drug lead screening technologies that are well-established to date, automation of the genome mining would drastically accelerate the drug development process (8).

1.2.2 Direct Cloning

Direct cloning was a widely used approach for BGC heterologous expression. Through simply cloning the whole BGC into an appropriate vector for its replication in the heterologous host, direct cloning appears to be one of the most widely used methods for heterologous expression. Although in most cases little changes are introduced to regulatory elements in the BGC, negative regulations can be bypassed, albeit not guaranteed by simply changing the *in vivo* environment for expression. While traditionally it is achieved by creating fosmid or cosmid libraries and screening for clones with target BGCs, which can be tedious and time consuming, direct cloning aims to specifically clone the target BGC by developing cutting-edge synthetic biology tools. Through integration of the recombineering sites to the target BGC, site specific recombineering systems such as Cre/loxP and phage φBT1 integrase system can separate the target BGC from the genome and circularize it as a plasmid (9, 10). By using the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 system, the target gene cluster can also be specifically cleaved from the genome. Jiang and co-workers developed a Cas9-assisted targeting of chromosome segments (CATCH) method for direct cloning of BGCs (11, 12). The target BGC is cleaved from intact genomic DNA by Cas9 *in vitro* and then ligated with a cloning vector through Gibson assembly. In addition to site specific integration or cleavage, the target BGC can also be selected from sheared genomic DNA fragments through transforming helper microorganisms with strong recombineering system, such as *E. coli* with the phage Rec/ET system or *Saccharomyces cerevisiae* (13, 14) (Figure 1.2).

Although direct cloning has been proven to be successful for bypassing the native regulatory system in many cases (15), its main limitation is still apparent for the heterologous expression of BGCs. As all putative regulatory elements in the BGCs are untouched in most cases, and the

regulatory mechanism for the expression of biosynthetic genes remain as a black box in most cases, the direct cloning strategies for successful expression of BGCs are more or less a trial and error process. The heterologous hosts are usually phylogenetically close to the native producers which provide similar environments for biosynthetic genes expression and enzymatic reactions, while the negative regulations which exist in the native producers are often missing in heterologous hosts. Even with continuous efforts for developing model hosts, the choices for expressing unique BGCs are still very limited, therefore lowering the successful rate for natural product discovery through direct cloning.

1.2.3 Pathway Refactoring

Pathway refactoring was used as an alternative strategy to direct cloning for natural product discovery through heterologous expression (3). By replacing the unknown regulatory elements with well characterized ones in the model hosts, the repression can be bypassed. Although pathway refactoring can be accomplished through stepwise ligation into the existing expression construct and co-transformation, it is a relatively low efficiency strategy and the number of genes which can be refactored is limited (16-18). Therefore, such a strategy typically requires a powerful DNA assembly method that can be used for reconstruction of the target BGC through assembly of multiple DNA fragments, including biosynthetic genes and regulatory elements, in a user-defined order (Figure 1.3). Considering that several tools developed for both *in vitro* (e.g. Gibson Assembly (19) and Golden Gate assembly (20)) and *in vivo* (e.g. DNA assembler (21)) DNA assembly can assemble multiple DNA fragments with high efficiency, there is great potential that such methods can be customized as standardized workflows for pathway refactoring and applied for natural product discovery (19, 20, 22).

One strategy to refactor a target BGC is through *in vivo* recombination, as demonstrated both in *E. coli* and *S. cerevisiae*. Apel and coworkers developed an artificial gene operon assembly system (AGOS) by using the Red/ET encoded *E. coli* (23). Biosynthetic genes or operons can be cloned into the entry plasmids harboring built-in promoters and terminators by restriction digestion-ligation. As all entry plasmids have different pairs of homology arms to the destination plasmid, different operons are integrated into the destination plasmid through recombination stepwisely in the Red/ET *E. coli* strain.

While the homologous recombination mechanism in *E. coli* has been used for pathway refactoring, the homologous recombination mechanism in *S. cerevisiae* has also been used for pathway refactoring in different ways. Brady and coworkers developed a method in which bidirectional *S. cerevisiae* promoters with different auxotrophic selection markers were built. In order to replace the native promoters in the BGC, homology arms to the target sites on the BGC are added to both ends of the bidirectional promoters and co-transformed with the gene cluster to *S. cerevisiae*. The successful replacement of native promoters can be isolated by auxotrophic selection. Such a method can insert up to three bidirectional promoters in one batch, if long homology arms (>500 bp) are used (24). Later on, they improved this method by integrating CRIPSR/Cas9 into this workflow and developed a yeast-based promoter engineering platform (mCRISTAR) (25). Due to the double strand breaks induced by CRIPSR/Cas9 during promoter integration, it has a significant increase in the recombination efficiency. Also by taking advantage of the highly efficient homologous recombination process in *S. cerevisiae*, Zhao and coworkers developed a method called DNA assembler where all DNA fragments, including regulatory parts and biosynthetic

genes, are prepared with homology arms to each other (21). After co-transformation into *S. cerevisiae*, all DNA fragments can be assembled into one construct. Through involving different parts for replication and selection, the final construct can be transferred into a different host for expression (26). By using this method, Zhao and coworkers successfully activated a silent polycyclic tetramate macrolactam (PTM) BGC from *S. griseus* in *S. lividans* and characterized the associated compounds (27).

Recently, a similar strategy has been applied for heterologous expression of fungus BGCs in *S. cerevisiae* through development of a heterologous expression (HEx) platform (28). Through pairing promoters and terminators as standard parts which serve as linkers for biosynthetic genes, fungal gene clusters can be simply refactored for expression in *S. cerevisiae* through transforming the biosynthetic genes and the corresponding standard parts as linear DNA into *S. cerevisiae* directly. To develop a suitable platform *S. cerevisiae* strain, they discovered and characterized pADH2-like promoters with various strength, which are auto-inducible when glucose and other fermentable carbon sources are converted to nonfermentable carbon sources. The delayed induction phenotype was confirmed by using these promoters for green fluorescence protein (GFP) expression. In addition, they developed an improved *S. cerevisiae* strain (DHY) by repairing the mitochondrial genome instability and deficiency of sporulation, deleting protease genes and integrating several genes for essential posttranslational modification enzymes (holo-ACP synthase and cytochrome P450 reductase). As a result, 22 new compounds were successfully detected from 41 refactored fungal BGCs.

In addition to *in vivo* DNA assembly methods, *in vitro* DNA assembly methods such as Gibson Assembly and Golden Gate assembly have also been used for pathway refactoring when used individually or in combination. By using Golden Gate assembly and Gibson Assembly together, Zhao and coworkers successfully refactored a 13-gene phosphonate BGC for heterologous expression in *S. lividans* and isolated a novel phosphonoacetic acid (29). In addition, Voigt and coworkers used Golden Gate assembly for refactoring silent BGCs such as the pyrrolnitrin BGC for expression in *E. coli* (30). Besides the activation of silent BGCs, through pathway refactoring with rational design, the performance of the pathway in heterologous hosts can be significantly improved. The nitrogen fixation gene cluster (nif) identified from *Klebsiella oxytoca* has undetectable activity when transformed into *E. coli* directly. However, after refactoring, 57% of the wild-type activity can be recovered in *E. coli* (30).

1.3 Ribosomally Synthesized and Post-translationally Modified Peptides (RiPPs)

Based on the structural and biosynthetic features of natural products discovered in the past century, natural products were traditionally divided into four categories: polyketides (PKs), nonribosomal peptides (NRPs), terpenoids and alkaloids. However, since the start of the 21st century, particularly due to the extensive application of advanced DNA sequencing technologies, another major category of natural products called ribosomally synthesized and post-translationally modified peptides (RiPPs) was identified. Unlike the NRP which requires the nonribosomal peptide synthase (NRPS) to build the skeleton of the final product, the biosynthesis of RiPPs starts from the expression of precursor peptides which commonly consist of two parts: leader peptide and core peptide. While post-translational modifications (PTMs) take place at the core peptide, the leader peptide serves as the recognition site for PTM enzymes. After all PTMs are finished, the leader

peptide should be cleaved off and the core peptide is transferred outside the cell to exert its function. To date, RiPPs with diverse PTMs have been discovered from all three domains of life and these RiPPs were also subdivided into more than 20 families (31). RiPPs also reveal diverse bioactivities such as antimicrobial, antifungal, and antiviral activities, making them an ideal resource for drug leads and some RiPPs are currently under clinical trials (32).

1.4 Polyketides

Polyketides are a major family of natural products with a broad range of bioactivities that have been studied for decades. The biosynthesis of polyketides starts from the formation of the polyketide skeleton which is catalyzed by polyketide syntheases (PKSs) following with various tailoring reactions such as oxidation and cyclization. PKSs assemble small acetic acid-type acyl building blocks, such as malonyl-CoA, into polyketides through C–C bonds. A common PKS consists of three types of core domains: acyltransferase (AT) domain, a ketosynthase (KS) domain and a carrier protein (CP) domain. The AT domain recognizes a specific acyl starter or extender unit and catalyzes its transfer onto the phosphopantetheine arm of the CP to form a thioester conjugate. The KS domain catalyzes a Claisen type condensation between an extender unit and a growing acyl chain. Sometimes the extending chain is also processed by ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains, which results in the formation of an α,β-double bond or a fully reduced methylene group. Polyketides have reveal diverse bioactivities including antimicrobial, antifungal and anticancer functions (33). For example, tetracycline is a widely used broad spectrum polyketide antibiotic produced by Streptomyces and daptomycin is another polyketide antibiotic which is usually used for life-threatening infections (34, 35). Epothilone is a 16-membered polyketide macrolactone and one of its analog ixabepilone has been

approved for cancer treatment (36). Therefore, it would be of great interest to discover novel polyketide natural products and study the biosynthetic mechanism.

1.5 Project Overview

My thesis research mainly focuses on developing and applying synthetic biology tools to accelerate natural product discovery. Throughout my work, I have developed a plug-and-play pathway refactoring workflow for natural product research. This workflow was then used for the discovery of glycocins and Class IV lanthipeptides, which reveal interesting structural features and bioactivities. In addition, *in vivo* and *in vitro* investigation of biosynthetic mechanisms were carried out for heat stable antifungal factor (HSAF) and a novel Class IV lanthipeptide respectively. Finally, I have obtained some preliminary data on the development of CRISPR-based approaches for gene regulation in Streptomycetes.

Chapter 2 describes the development of a plug-and-play pathway refactoring workflow for high-throughput, flexible pathway construction, and expression in both *E. coli* and *S. cerevisiae*. Biosynthetic genes were firstly cloned into pre-assembled helper plasmids with promoters and terminators, resulting in a series of expression cassettes. These expression cassettes were further assembled using Golden Gate reaction to generate fully refactored pathways. The inclusion of spacer plasmids in this system would not only increase the flexibility for refactoring pathways with different number of genes, but also facilitate gene deletion and replacement. As a proof of concept, the zeaxanthin biosynthetic pathway was refactored with more than 95% fidelity and successfully expressed in *S. cerevisiae*. Moreover, in order to further prove the flexibility of this workflow for generating gene knockouts, another three zeaxanthin pathways with different gene

knockouts were also constructed and the corresponding intermediates were isolated after expression in *S. cerevisiae*. This workflow was also proven to be useful for combinatorial biosynthesis by refactoring another three carotenoid pathways with different gene combinations for expression in *E. coli*. To demonstrate its application for high-throughput pathway refactoring, a total of 96 pathways for combinatorial carotenoid biosynthesis were built successfully.

Chapter 3 describes the application of the pathway refactoring workflow in Chapter 2 for discovering glycocins (glycosylated bacteriocins). Glycocins are a family of RiPPs with antimicrobial activities against pathogens of interest, including methicillin–resistant Staphylococcus aureus, representing a promising source of new antibiotics. Glycocins are still largely underexplored, and thus far, only six glycocins are known. In this chapter, 50 putative glycocin biosynthetic gene clusters were identified by genome mining and six of them with distinct features were chosen for further investigation. Through two rounds of plug-and-play pathway refactoring and expression in *E. coli* BL21(DE3), four systems produced novel glycocins. Further structural characterization revealed that one of them, which belongs to the enterocin 96-type glycocins, was diglucosylated on a single serine. The other three compounds belong to the SunA/ThuA-type glycocins and exhibit an antimicrobial spectrum narrower than that of sublancin, the best characterized member in this group, even though they share a similar disulfide topology and glycosylation. Further evaluation of their bioactivities with free glucose at high concentrations suggested that their antimicrobial mechanisms might be both glycocin- and species-specific. These glycocins with distinct features significantly broaden our knowledge and may lead to the discovery of new classes of antibiotics.

In Chapter 4, a novel Class IV lanthipeptide was discovered by the same method as I described in Chapter 3. Lanthipeptides (lanthionine-containing peptides) constitute a major family RiPPs which have thioether crosslinks as their signature structural feature. Although most lanthipeptides exhibit antimicrobial bioactivity, other bioactivities such as antifungal (37), morphogenetic (38, 39), antiviral (40), antinociceptive (41), and antiallodynic functions (42) were also discovered. Based on the characteristics of lanthipeptide synthases, lanthipeptides were subdivided into four classes (Class I-Class IV). While extensive studies have been carried out for Class I and Class II lanthipeptides, there is little knowledge for Class III and Class IV lanthipeptides. Particularly, only a few Class IV lanthipeptides have been characterized and they all structurally resemble to venezuelin, a prototype Class IV lanthipeptide. In this chapter, I successfully identified a novel Class IV lanthipeptide with completely new structure. To further characterize its biosynthesis, the post-translational modification was reconstituted *in vitro*. By analyzing the structure of the intermediates of biosynthesis, the lanthipeptide synthase reveal a bi-directional cyclization activity.

Chapter 5 describes the characterization of the biosynthesis of heat stable antifungal factor (HSAF), which is a polyketide belonging to polycyclic tetramate macrolactam. Previously Zhao and coworkers activated a cryptic polycyclic tetramate macrolactam pathway from *Streptomyces griseus* and proposed a parallel biosynthetic mechanism for this pathway. To further investigate if such a parallel biosynthetic mechanism generally exists for other polycyclic tetramate macrolactams, I chose the HSAF pathway and carried out multiple gene knockouts in order to identify the key intermediates. Three intermediates were successfully isolated and characterized by mass spectrometry and NMR, which supports the assumption that the HSAF pathway also has

a parallel biosynthetic mechanism. Bioinformatic studies also reveal that such a biosynthetic mechanism might generally exist in most of the polycyclic tetramate macrolactam pathways.

Chapter 6 describes some preliminary data about the development of CRISPR-based approaches for titratable activation and repression of genes in Streptomyces species. Streptomyces are ideal chassis for discovery and overproduction of medically relevant secondary metabolites (43). However, the shortage of genetic engineering tools in streptomyces largely restricted their practical applications. In this chapter, I demonstrated the activity of CRISPR-dCas9 in *Streptomyces lividans* 66 and *Streptomyces albus* J1074. By attaching the RNA polymerase omega factor (RNAP-ω) to the C-terminal of Cas9 protein, the fusion protein was able to increase the gene expression level through targeting the corresponding promoter region. Of note, the titratable repression of target genes was supposed to be achieved by using an inducible promoter to tune the expression of gRNA. However, the leakage of gRNA expression led to significant target repression without any induction and thus a limited tunable range. Therefore, systematic optimization to fine-tune the expression of both gRNA and dCas9 is required.

1.6 Figures

Figure 1.1. Overview of the workflow for a modern natural product discovery strategy. BGCs are first identified from genomic sequences by bioinformatics tools and then activated by different synthetic biology strategies. Products are characterized by high-throughput methods and the data obtained can further guide the next round genome mining.

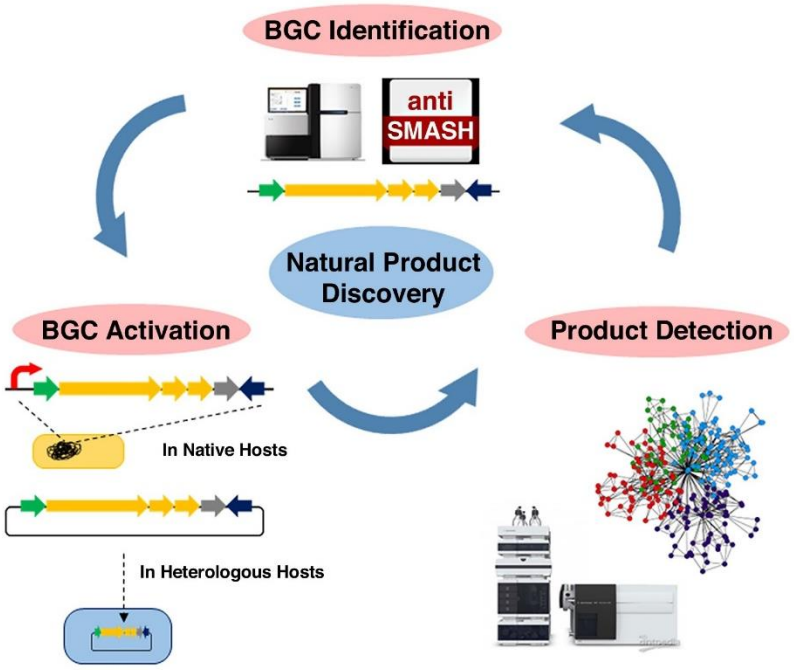


Figure 1.2. Representative methods for direct cloning of BGCs. (a) TAR and RecET mediated in vivo recombination; (b) Cre/loxP and phage φBT1 mediated site-specific recombination; (c) CATCH.

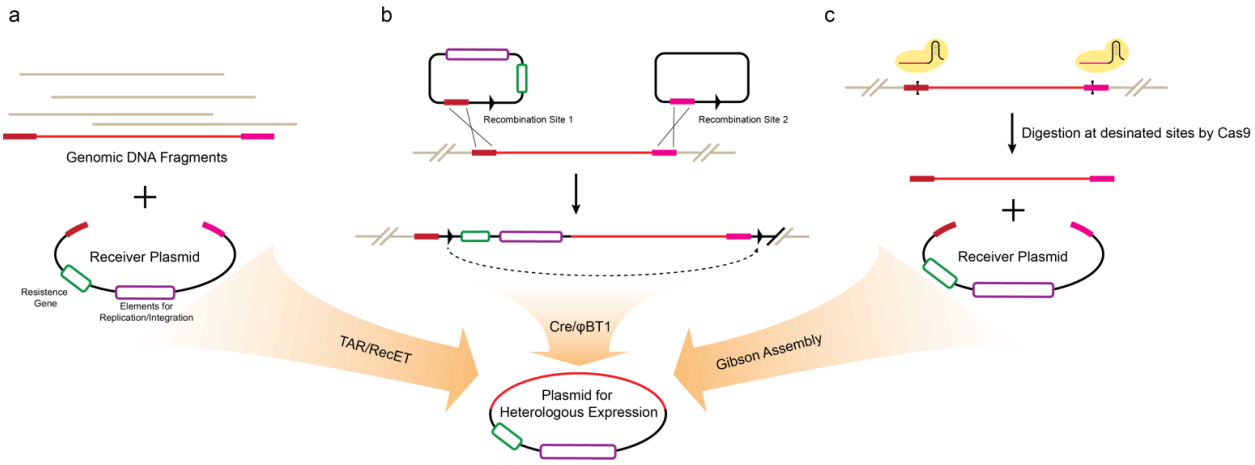
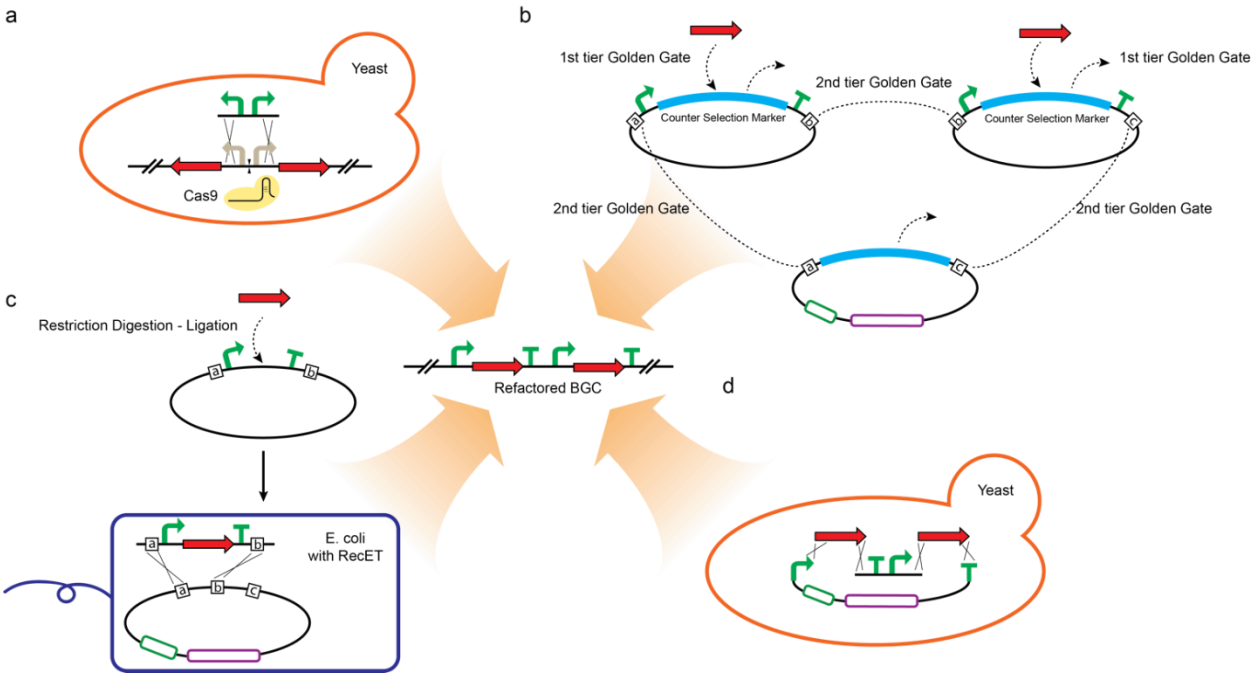


Figure 1.3. Representative methods for refactoring of BGCs. (a) mCRISTAR; (b) Plug-and-play Golden Gate assembly based pathway refactoring; (c) Artificial gene operon assembly system (AGOS); (d) DNA assembler.



1.7 References

1. Newman DJ & Cragg GM (2016) Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod* 79(3):629-661.
2. Demain AL & Sanchez S (2009) Microbial drug discovery: 80 years of progress. *J Antibiot* 62(1):5-16.
3. Rutledge PJ & Challis GL (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* 13(8):509-523.
4. Li JWH & Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? *Science* 325(5937):161-165.
5. Bentley SD, *et al.* (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* 417(6885):141-147.
6. Ikeda H, *et al.* (2003) Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nat Biotechnol* 21(5):526-531.
7. Ren HQ, Wang B, & Zhao HM (2017) Breaking the silence: new strategies for discovering novel natural products. *Curr Opin Biotech* 48:21-27.
8. Smanski MJ, *et al.* (2016) Synthetic biology to access and expand nature's chemical diversity. *Nat Rev Microbiol* 14(3):135-149.
9. Hu SB, *et al.* (2016) "Cre/loxP plus BAC": a strategy for direct cloning of large DNA fragment and its applications in *Photorhabdus luminescens* and *Agrobacterium tumefaciens*. *Sci Rep* 6:29087.
10. Dai RX, Zhang B, Zhao GP, & Ding XM (2015) Site-specific recombination for cloning of large DNA fragments *in vitro*. *Eng Life Sci* 15(6):655-659.
11. Jiang WJ & Zhu TF (2016) Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat Protoc* 11(5):960-975.
12. Jiang WJ, *et al.* (2015) Cas9-assisted targeting of chromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun* 6.
13. Fu J, *et al.* (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol* 30(5):440.
14. Yamanaka K, *et al.* (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *P Natl Acad Sci USA* 111(5):1957-1962.
15. Li L, Jiang WH, & Lu YH (2017) New strategies and approaches for engineering biosynthetic gene clusters of microbial natural products. *Biotechnol Adv* 35(8):936-949.
16. Chang FY, Ternei MA, Calle PY, & Brady SF (2013) Discovery and synthetic refactoring of tryptophan dimer gene clusters from the environment. *J Am Chem Soc* 135(47):17906-17912.
17. Kuthning A, Mosker E, & Sussmuth RD (2015) Engineering the heterologous expression of lanthipeptides in *Escherichia coli* by multigene assembly. *Appl Microbiol Biot* 99(15):6351-6361.
18. McClerren AL, *et al.* (2006) Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. *P Natl Acad Sci USA* 103(46):17243-17248.
19. Gibson DG, *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343-U341.

20. Engler C, Gruetzner R, Kandzia R, & Marillonnet S (2009) Golden Gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *Plos One* 4(5):e5553.
21. Shao Z, Zhao H, & Zhao H (2009) DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* 37(2):e16.
22. Shao ZY & Zhao HM (2012) DNA assembler: a synthetic biology tool for characterizing and engineering natural product gene clusters. *Method Enzymol* 517:203-224.
23. Basitta P, *et al.* (2017) AGOS: a plug-and-play method for the assembly of artificial gene operons into functional biosynthetic gene clusters. *ACS Synth Biol* 6(5):817-825.
24. Montiel D, Kang HS, Chang FY, Charlop-Powers Z, & Brady SF (2015) Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *P Natl Acad Sci USA* 112(29):8953-8958.
25. Kang HS, Charlop-Powers Z, & Brady SF (2016) Multiplexed CRISPR/Cas9-and TAR-mediated promoter engineering of natural product biosynthetic gene clusters in yeast. *ACS Synth Biol* 5(9):1002-1010.
26. Shao ZY & Zhao HM (2013) Construction and engineering of large biochemical pathways via DNA assembler. *Methods Mol Biol* 1073:85-106.
27. Luo YZ, *et al.* (2013) Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nat Commun* 4.
28. Harvey CJB, *et al.* (2018) HEx: A heterologous expression platform for the discovery of fungal natural products. *Sci Adv* 4(4):eaar5459.
29. Freestone TS, Ju KS, Wang B, & Zhao HM (2017) Discovery of a phosphonoacetic acid derived natural product by pathway refactoring. *ACS Synth Biol* 6(2):217-223.
30. Smanski MJ, *et al.* (2014) Functional optimization of gene clusters by combinatorial design and assembly. *Nat Biotechnol* 32(12):1241-U1104.
31. Arnison PG, *et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30(12):1568-1568.
32. Dang T & Suesmuth RD (2017) Bioactive peptide natural products as lead structures for medicinal use. *Accounts Chem Res* 50(7):1566-1576.
33. Gomes ES, Schuch V, & Lemos EGD (2013) Biotechnology of polyketides: New breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Braz J Microbiol* 44(4):1007-1034.
34. Steenbergen JN, Alder J, Thorne GM, & Tally FP (2005) Daptomycin: a lipopeptide antibiotic for the treatment of serious Gram-positive infections. *J Antimicrob Chemoth* 55(3):283-288.
35. Chopra I & Roberts M (2001) Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol R* 65(2):232-260.
36. Demain AL & Vaishnav P (2011) Natural products for cancer chemotherapy. *Microb Biotechnol* 4(6):687-699.
37. Mohr KI, *et al.* (2015) Pinensins: the first antifungal lantibiotics. *Angew Chem Int Edit* 54(38):11254-11258.
38. Kodani S, *et al.* (2004) The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*. *P Natl Acad Sci USA* 101(31):11448-11453.

39. Kodani S, Lodato MA, Durrant MC, Picart F, & Willey JM (2005) SapT, a lanthionine-containing peptide involved in aerial hyphae formation in the streptomycetes. *Mol Microbiol* 58(5):1368-1380.
40. Ferir G, *et al.* (2013) The Lantibiotic peptide labyrinthopeptin A1 demonstrates broad anti-HIV and anti-HSV activity with potential for microbicidal applications. *Plos One* 8(5).
41. Iorio M, *et al.* (2014) A glycosylated, labionin-containing lanthipeptide with marked antinociceptive activity. *ACS Chem Biol* 9(2):398-404.
42. Meindl K, *et al.* (2010) Labyrinthopeptins: a new class of carbacyclic lantibiotics. *Angew Chem Int Edit* 49(6):1151-1154.
43. Liu R, Deng ZX, & Liu TG (2018) Streptomyces species: ideal chassis for natural product discovery and overproduction. *Metab Eng* 50:74-84.

Chapter 2: A Plug-and-Play Pathway Refactoring Workflow for Natural Product Research in *Escherichia coli* and *Saccharomyces cerevisiae*

2.1 Introduction

Natural products are an indispensable source for drug discovery (1). However, due to the ever-increasing rediscovery rate of natural products by traditional methods, an important approach to expand the natural product library for screening new drug candidates is combinatorial biosynthesis (2). Thanks to the great efforts on natural product research in the past decades, biosynthetic genes with various functions have been continuously characterized. Moreover, the advances in next generation sequencing and bioinformatics have enabled rapid prediction of numerous biosynthetic gene clusters (BGCs) with the potential to generate novel natural products (3, 4). Nevertheless, most studies in combinatorial biosynthesis are limited to only a small fraction of all possible combinations, mainly due to the lack of a high-throughput pathway refactoring method (5). Besides combinatorial biosynthesis, another intriguing part of natural product research is the investigation of biosynthetic mechanisms, which usually requires gene deletion in a BGC. However, gene deletion by constructing pathway variants through existing DNA assembly methods usually affects the context of other genes, resulting in repetitive cloning work. Therefore, it is highly desirable to develop a fast, high-throughput and flexible pathway refactoring method for natural product research.

Pathway refactoring can be technically accomplished through DNA assembly. Golden Gate reaction is a DNA assembly method based on the activity of Type IIs restriction enzymes, which cut outside their recognition sites and generate single-strand DNA overhangs (6). When designed appropriately, the overhangs can guide the corresponding DNA fragments to be ligated in a

designated order by a DNA ligase. Previously, we optimized the condition of Golden Gate assembly with high and consistent fidelity (above 95%) and successfully applied it for high-throughput synthesis of transcription activator-like effectors (TALEs) (7). In this work, we extended this strategy to establish a scalable pathway refactoring workflow. Briefly, multiple helper plasmids are constructed with well characterized promoters and terminators between which biosynthetic genes can be inserted. The constructed cassettes in the helper plasmid can then be assembled in a defined order with the help of 4 bp overhangs. Spacer plasmids, which share the same 4 bp overhangs with their corresponding helper plasmids but only encode a 20 bp random DNA sequence are also constructed and help the refactoring workflow adapt to pathways with different number of genes. We employed this workflow to refactor the zeaxanthin biosynthetic pathway and achieved high fidelity. Because of the highly modularized design, the gene deletion and replacement can also be realized by appropriately using the spacer plasmids. As proof of concept, 96 functional pathways for combinatorial carotenoid biosynthesis were successfully built using this workflow.

2.2 Results and Discussion

2.2.1 Design of the Plug-and-Play Pathway Refactoring Workflow

As shown in Figure 2.1, the workflow consists of two tiers of Golden Gate reaction, which are catalyzed by *Bbs*I (1st tier) and *Bsa*I (2nd tier) respectively. The biosynthetic gene can be either synthesized or PCR amplified with two *Bbs*I cleavage sites at both ends, which help in generating the general overhangs AATG (at the side of start codon) and CGGT (at the side of stop codon) respectively. Other *Bbs*I and *Bsa*I cleavage sites in the biosynthetic gene need to be removed at the same time through silent mutations. The same *Bbs*I cleavage sites and the corresponding

overhangs also exist on the helper plasmid, flanking a counter-selection marker (*ccdB*). Those *Bbs*I cleavage sites on the helper plasmid are also flanked by a promoter and a terminator (Figure 2.1A). In the 1st tier *Bbs*I catalyzed Golden Gate reaction, the *ccdB* marker on the helper plasmid is replaced by the biosynthetic gene, resulting in an expression cassette. In the overhang AATG between the promoter and the biosynthetic gene, the first "A" comes from the last nucleotide of the promoter followed by the start codon "ATG" of the biosynthetic gene. This design allows the seamless connection of the promoter and the biosynthetic gene.

After the 1st tier reaction, all the helper plasmids containing expression cassettes can be ligated together by their unique overhangs generated by *Bsa*I in the 2nd tier reaction. However, as the number of biosynthetic genes varies from one BGC to another, gaps can appear once the number of biosynthetic genes is less than the number of helper plasmids (Figure 2.1B). Therefore, another set of plasmids called "spacer plasmids" are included in the system which have the same overhangs with their corresponding helper plasmids, but only a random 20 bp DNA sequence in the middle. When the helper plasmid is not used in the 1st tier reaction, the corresponding spacer plasmid can be used to "fill the gap" in that position (Figure 2.1B). As a result, the 4 bp overhangs (ATGG and AGCG) flanking the *ccdB* marker on the receiver plasmid are not affected and the same receiver plasmid can be used for assembling pathways with various number of expression cassettes.

2.2.2 Refactoring of the Zeaxanthin Biosynthetic Pathway Using *S. cerevisiae* Promoters and Terminators

As proof of concept, we applied this workflow to refactor the zeaxanthin biosynthetic pathway in *S. cerevisiae* (8). Nine helper plasmids were built using the promoters and terminators from *S.*

cerevisiae and the corresponding spacer plasmids were also built with the 20 bp random sequences designed by R2oDNA (Table 1) (9). We first checked the fidelity of the 1st tier reaction by blue-white screening. The *lacZ* cassette was cloned into one of the helper plasmids and the reaction mixture was transformed into NEB10-beta and spread on an LB+Amp plate with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) and isopropyl β-D-1-thiogalactopyranoside (IPTG). After overnight incubation, all the colonies on plate were blue, indicating the fidelity of the 1st tier reaction was 100% (Figure 2.2). By using the same reaction condition, all of the five genes from the zeaxanthin pathway were cloned into different *S. cerevisiae* helper plasmids respectively and 6 colonies were picked from each reaction. The plasmids isolated from these clones were digested by *Bsa*I and all of them showed the correct digestion patterns (Figure 2.3A). Then the 5 plasmids with the corresponding expression cassettes were mixed with the other 4 spacer plasmids and the receiver plasmid for the 2nd tier Golden Gate reaction. After transformation into NEB10-beta and overnight incubation, constructs isolated from all 20 transformants showed the expected digestion patterns, indicating 100% assembly fidelity (Figure 2.3C).

Since both the 1st and 2nd tier Golden Gate reactions showed very high fidelity, we hypothesized that polyclonal plasmids could also be used for cloning and expression, which would save the time for colony growth on the plate and the labor for colony picking. As a polyclonal plasmid can be obtained from either the 1st or 2nd tier reaction, there are four scenarios to obtain the final construct (I: monoclonal-monoclonal; II: monoclonal-polyclonal; III: polyclonal-monoclonal; IV: polyclonal-polyclonal). Polyclonal plasmids from the 1st tier reaction were digested and showed no significant difference from their monoclonal counterparts (Figure 2.3B), and these plasmids were used for 2nd tier reaction for scenarios III and IV. In scenario III, 20 colonies from the 2nd tier

reaction were picked for plasmid isolation and digestion, 19 out of 20 showed a correct pattern (Figure 2.3D). In scenarios II and IV, polyclonal plasmids from the 2nd tier reaction were also checked by restriction digestion and no significant difference from the monoclonal one can be observed (Figure 2.3E). The final constructs from scenario I to scenario IV were transformed into *S. cerevisiae* CEN.PK2-1C for expression. Cells were extracted by acetone and analyzed by HPLC. All samples showed peaks at 430 nm with the same retention time as the zeaxanthin standard, which confirmed the production of zeaxanthin (Figure 2.3F-G). By using polyclonal cultures, the workflow can be finished in two days, which is suitable for quick check of the pathway's function. However, we recommend using monoclonal cultures for quantitative analysis of the pathway's function, which only requires some additional waiting time (1 or 2 overnight) for colony growth.

2.2.3 Construction of Various Zeaxanthin Pathways with Gene Deletions Using Spacer Plasmids

Because of its high modularity, this workflow also enables researchers to delete genes in the final construct for biosynthetic mechanistic studies with minimal cloning effort. Because every expression cassette should have the same overhangs that are generated by *Bsa*I with the corresponding spacer plasmid, inclusion/deletion of a gene can be readily accomplished by using either the expression cassette or the corresponding spacer plasmid in the 2nd tier reaction and no repetitive cloning effort in the 1st tier is needed. As proof of concept, the same expression cassettes for the construction of the zeaxanthin pathway were used to build pathways producing phytoene, lycopene and β-carotene, which are intermediates in the zeaxanthin biosynthesis (10). The final constructs were again transformed into CEN.PK2-1C and extracted by acetone. The expected colors associated with these products were observed for all the three samples (Figure 2.4C). The

samples were further analyzed with HPLC and LC/MS and the production of phytoene **1**, lycopene **2** and β-carotene **3** was verified (Figure 2.4 and Figure 2.6).

2.2.4 Construction of Pathways for Combinatorial Carotenoid Biosynthesis in *E. coli*

Because the overhangs generated by *Bsa*I are independent from the sequence of the biosynthetic gene, every gene in the final construct can be easily replaced by another gene sharing the same helper plasmid. As a demonstration, two C30 carotenoid pathways with only one gene difference were selected (11) (Figure 2.5A). CrtM and CrtN produce the intermediate diapolycopene shared by those two pathways and then CrtA or CrtOx adds a ketone group (or a hydroxyl group) at different positions (12, 13). As all the four enzymes showed little activity in *S. cerevisiae*, another set of helper plasmids with T7 promoters and terminators (four in total) was built for the expression in *E. coli*. The final constructs were transformed into BL21(DE3) for expression and the yellow, orange and red colors were observed in the corresponding crude extracts (Figure 2.5C). The crude extracts were again analyzed by the HPLC and LC/MS. The diapolycopene **4** was observed from CrtM+CrtN at 430 nm. The expected product compounds **5** and **6** were verified from CrtM+CrtN+CrtA by LC/MS while compounds **7**, **8** and **9** were verified from CrtM+CrtN+CrtOx by LC/MS as well (Figure 2.5 and Figure 2.6).

Finally, to investigate whether the plug-and-play workflow has the potential to be scaled up and applied in combinatorial biosynthesis, a total of 96 pathways were created for the biosynthesis of carotenoids. Carotenoid biosynthesis can be split into three steps: a) backbone synthesis, b) desaturation, and c) cyclization or other modification (11, 14). Taking the biosynthesis of β-carotene as an example, the CrtE and CrtB first catalyze the formation of the C40 backbone

phytoene, then the double bonds are introduced by CrtI, and finally the two ring structures are formed by CrtY. Accordingly, we collected 15 carotenoid biosynthetic genes and mutants and divided them into 3 groups. The backbone synthases group (group 1) includes four subgroups which can produce carotenoid backbones ranging from C30 to C50 (15, 16). The desaturases group (group 2) contains wild type C40 desaturase CrtI and its two mutants and wild type C30 desaturase CrtN (10). The cyclases and others group (group 3) consists of six enzymes which can catalyze different cyclization and oxidation reactions (10, 17) (Figure 2.7A). As the longest pathway consists of four genes (two from group 1, one from group 2 and one from group 3), the 4 helper plasmids that we built for the expression in *E. coli* should be sufficient for constructing all of the 96 combinations. Genes from group 1 were assigned to either helper plasmid 1 or helper plasmid 2, while genes from group 2 and 3 were assigned to helper plasmids 3 and 4 respectively. Successful assembly of the 96 constructs were confirmed using restriction digestion, and the plasmids were transformed into BL21(DE3) for carotenoid production. Carotenoids with different colors were generated after overnight induction by IPTG (Figure 2.7B).

2.3 Conclusion

In this work we have developed a pathway refactoring workflow based on hierarchical Golden Gate assembly. Two sets of helper plasmids and one set of spacer plasmids with standardized 4 bp overhangs were constructed, which enabled the addition, deletion and exchange of expression cassettes in the final construct with minimal cloning effort. By using the optimized condition of Golden Gate assembly, high-throughput pathway refactoring was accomplished and demonstrated by the successful construction of 96 combinatorial pathways for carotenoid biosynthesis.

Prior to this work, a few Golden Gate reaction based multigene assembly systems have been developed, such as the GoldenBraid and the modular cloning (MoClo) system (18, 19). Among all these systems, the most representative one is the MoClo and currently there are multiple toolkits and part libraries available to enable its application in various hosts, such as *E. coli* and *S. cerevisiae* (20, 21). By its delicate design, MoClo enables researchers to assemble multigene constructs with very high flexibility, as all the parts, such as promoters and terminators, are modularized and free to be changed. But in the natural product research, since the discovery of new natural products and the study of biosynthetic mechanisms are usually the focuses, it is unnecessary to find out the best combination of promoters for the BGC and in most cases strong promoters are arbitrarily chosen. Therefore, we include all the elements that are necessary for gene expression in our helper plasmids. Since only the target gene and the helper plasmid are needed for the 1st tier reaction, potentially repetitive work to add the same parts, such as promoters and terminators, are avoided, which makes the process friendlier for high-throughput automation-based natural product researches, even though the flexibility is more or less affected. In addition, MoClo has also designed many end-linkers and destination vectors to help its users add or delete genes at both ends in their constructs, but it is unable to delete the gene in the middle without including the selectable marker. However, in the study of biosynthetic mechanisms, researchers usually do not know the gene functions in the BGC. As a result, genes at any positions of their construct might need to be deleted. Thus, we include the spacer plasmids in our workflow, which not only makes the gene deletion become easier, but also reduces the total number of the working plasmids (helper plasmids, receiver plasmids, etc.) that are needed in the system. Overall, our workflow is more focused on natural product research and simplified to make it more automation friendly.

In addition to the applications for biosynthetic mechanism investigation and combinatorial biosynthesis as demonstrated in this work, another potential application of this workflow would be the activation of silent BGCs (22), which is currently underway in our laboratory. In summary, this pathway refactoring workflow can serve as a powerful tool for the natural product research in the future.

2.4 Materials and Methods

2.4.1 Strains, Media and Cell Cultivation

All chemicals were purchased from Sigma-Aldrich or ThermoFisher Scientific. *E. coli* NEB® 10-beta (New England Biolabs, Ipswich, MA) was used for DNA manipulation. Plasmids containing *ccdB* marker are propagated in *E. coli* DB3.1 (*gyrA462 endA1 Δ(sr1-recA) mcrB mrr hsdS20 glnV44 (=supE44) ara14 galK2 lacY1 proA2 rpsL20 xyl5 leuB6 mtl1*). *S. cerevisiae* strain CEN.PK2-1C (*MATa ura3−52 trp1−289 leu2−3,112 his3Δ1 MAL2−8C SUC2*) (EUROSCARF) and *E. coli* BL21(DE3) (Cell Media Facility, University of Illinois at Urbana−Champaign, Urbana, IL) were used for carotenoid production. *E. coli* was cultured in Lysogeny broth (LB) (ThermoFisher Scientific, Waltham, MA) containing appropriate antibiotics at 37 °C and 250 rpm unless specified otherwise. *S. cerevisiae* was cultured in synthetic complete medium (SC-URA) consisting of 0.17% Difco yeast nitrogen base without amino acids, 0.5% ammonium sulfate, and 0.083% amino acid drop out mixture without uracil (MP Biomedicals, Solon, OH) supplemented with 2% glucose was used for DNA homologous recombination and zeaxanthin fermentation. The plasmids containing *S. cerevisiae* promoters and terminators and the genes (*crtE, crtB, crtI, crtY, crtZ*) were described elsewhere (8). The plasmids containing other carotenoid biosynthetic genes

were gifts from Frances Arnold at the California Institute of Technology and Claudia Schmidt-Dannert at the University of Minnesota.

2.4.2 1st Tier Golden Gate Reaction

100 ng helper plasmid was mixed with an insert in 1:3 molar ratio. The reaction condition was as follows: a 20 μL reaction system contained 2 μL 10x Tango buffer (ThermoFisher Scientific, Waltham, MA), 1 μL 10 mM dithiothreitol (DTT), 1 μL 10 mM adenosine triphosphate (ATP), 1 μL FastDigest *Bbs*I (ThermoFisher Scientific, Waltham, MA), 0.5 μL T7 ligase (Enzymatics, Beverley, MA), and ddH_2O to 20 μL. The following temperature cycles were set up: 6 cycles of 5 min at 37 ℃ and 5 min at 21 ℃. Then the reaction mix was treated by Plasmid-Safe-ATP-Dependent DNase (Epicentre, Madison, WI) by adding 0.25 μL enzyme and 1 μL 25 mM ATP and then incubating at 37 ℃ for 30 min. 10 μL reaction mix was then transformed into 100 μL NEB® 10-beta competent cell by standard heat-shock transformation. After recovery in 1 mL LB at 37 ℃ for 1 h, 5 μL was inoculated into 5 mL LB with 100 μg/mL ampicillin for polyclonal plasmid isolation or spread on LB+Amp plate and incubate overnight at 37 ℃. Single colony was inoculated into 5 mL LB liquid culture with 100 μg/mL ampicillin next day for monoclonal plasmid isolation. Plasmids were isolated using QIAGEN Plasmid Mini Kit (QIAGEN, Germantown, MD). The maps of the helper plasmids used in this work are shown in Figure 2.8 and Figure 2.9.

2.4.3 2nd Tier Golden Gate Reaction

100 ng receiver plasmid (~0.035 nmol) was mixed with the same molar amount of helper plasmids with corresponding inserts cloned in and the appropriate spacer plasmids. The reaction condition

was as follows: a 20 μL reaction system contained 2 μL 10x T4 DNA Ligase Reaction Buffer (New England Biolabs, Ipswich, MA), 0.8 μL *Bsa*I (New England Biolabs, Ipswich, MA), 0.2 μL T4 DNA Ligase (M0202T, New England Biolabs, Ipswich, MA), and ddH_2O to 20 μL. The following temperature cycles were set up: 5 min at 37 ℃, 30 cycles of 5 min at 37 ℃ and 10 min at 16 ℃, 30 min at 16 ℃, 45 min at 37 ℃ and 5 min at 80 ℃. The reaction was then treated by Plasmid-Safe-ATP-Dependent DNase (Epicentre, Madison, WI) and transformed into 100 μL NEB® 10-beta competent cell, which was similar to the 1st tier Golden Gate reaction. Monoclonal and polyclonal plasmids were isolated by the same method that described in the 1st tier Golden Gate reaction. Plasmids were isolated using QIAGEN Plasmid Mini Kit (QIAGEN, Germantown, MD). The maps and sequences of the spacer and receiver plasmids used in this work are available in Figures 2.10-2.12.

2.4.4 Design of the Spacer Plasmid Inserts

All insert sequences were designed using R2oDNA (http://www.r2odna.com/) with the following criteria: GC content of 60%; no start codons, common multiple cloning sites (MCS), or assembly restriction sites; no bacterial promoter-like sequences; and no similar sequences within *E. coli* and *S. cerevisiae* genome. Detailed sequences of 18 unique inserts utilized in this study are listed in Table 1.

2.4.5 Functional Analysis of the Assembled Pathways

For carotenoid pathways expressed in *S. cerevisiae*, cell pellets were harvested from either 50 mL (for analysis of zeaxanthin) or 500 mL (for analysis of phytoene, lycopene and β-carotene) Sc-Ura liquid media after fermentation at 30 ℃ and 250 rpm for 72 h. The extraction was performed as

previously described (8). Briefly the cell pellet was resuspended by 50 mL acetone and passed through French press at 10,000 psi three times.

Pathways for carotenoid biosynthesis in *E. coli* was induced by 0.1 mM IPTG when the OD600 reached 0.6-0.8 and the cell pellets containing the CrtM+CrtN, CrtM+CrtN+CrtA, or CrtM+CrtN+CrtOx pathways were harvested from 500 mL LB liquid media after fermentation at 30 ℃ and 250 rpm for 24 h. The cell pellet was resuspended in 50 mL acetone and vortexed for 1 min.

After extraction from either *S. cerevisiae* or *E. coli*, the supernatant was collected after centrifugation and evaporated to dryness. After resuspension in 200 μL THF, 20 μL was analyzed on an Agilent 1100 series HPLC with a Photodiode Array (PDA) detector. H_2O + 0.1% formic acid (solvent A) and acetonitrile + 0.1% formic acid (solvent B) were used as mobile phase and different reverse-phase columns were used for different samples. The zeaxanthin containing crude extract was loaded onto Kinetex® 2.6 μm EVO C18 100Å 150x2.1 mm column (Phenomenex, Torrance, CA) and eluted at 0.2 mL/min with the following gradient: 50% B for 5 min, to 100% B in 10 min, hold at 100% B for 10 min and return to 50% B over 10 min. The other samples were loaded onto Kinetex® 5 μm EVO C18 100Å 250x4.6 mm column (Phenomenex, Torrance, CA) and eluted at 1 mL/min with the following gradient: 70% B for 15 min, to 100% B in 60 min, hold at 100% B for 30 min and return to 70% over 15 min. The same condition was used for the analysis of carotenoids by LC/MS (Agilent). Mass spectra were acquired in ultra scan mode using atmospheric pressure chemical ionization with positive polarity. The MS system (Agilent 1100

LC/MSD Trap XCT Plus) was operated using a drying temperature of 350 °C, a nebulizer pressure of 35 psi, a drying gas flow of 8.5 L/min, and a capillary voltage of 4500 V.

The 96 combinatorial pathways for carotenoid biosynthesis in *E. coli* were similarly induced by 0.1 mM IPTG and fermented at 30 ℃ and 250 rpm for 24 h. Then 1 mL culture was taken from 5 mL culture for each pathway and placed into the corresponding position in a 96 deep well plate. Cell pellets were harvested by centrifugation.

2.5 Figures and Table

Figure 2.1. Scheme of the plug-and-play pathway refactoring workflow. (A) The 1st tier Golden Gate reaction. The gene is either synthesized or PCR amplified with *Bbs*I cleavage sites at both ends and cloned into a helper plasmid through *Bbs*I catalyzed Golden Gate reaction. (B) The 2nd tier Golden Gate reaction. Helper plasmids harboring the corresponding genes are mixed with the appropriate spacer plasmids and the receiver plasmid and assembled into the final construct through *Bsa*I catalyzed Golden Gate reaction. All helper plasmids and spacer plasmids share the pUC19 backbone, while the receiver plasmid has either a pET28a backbone (for expression in *E. coli*) or pRS416 backbone with the ampicillin resistance gene replaced by the kanamycin resistance gene (for expression in *S. cerevisiae*). This figure was adapted from (23).

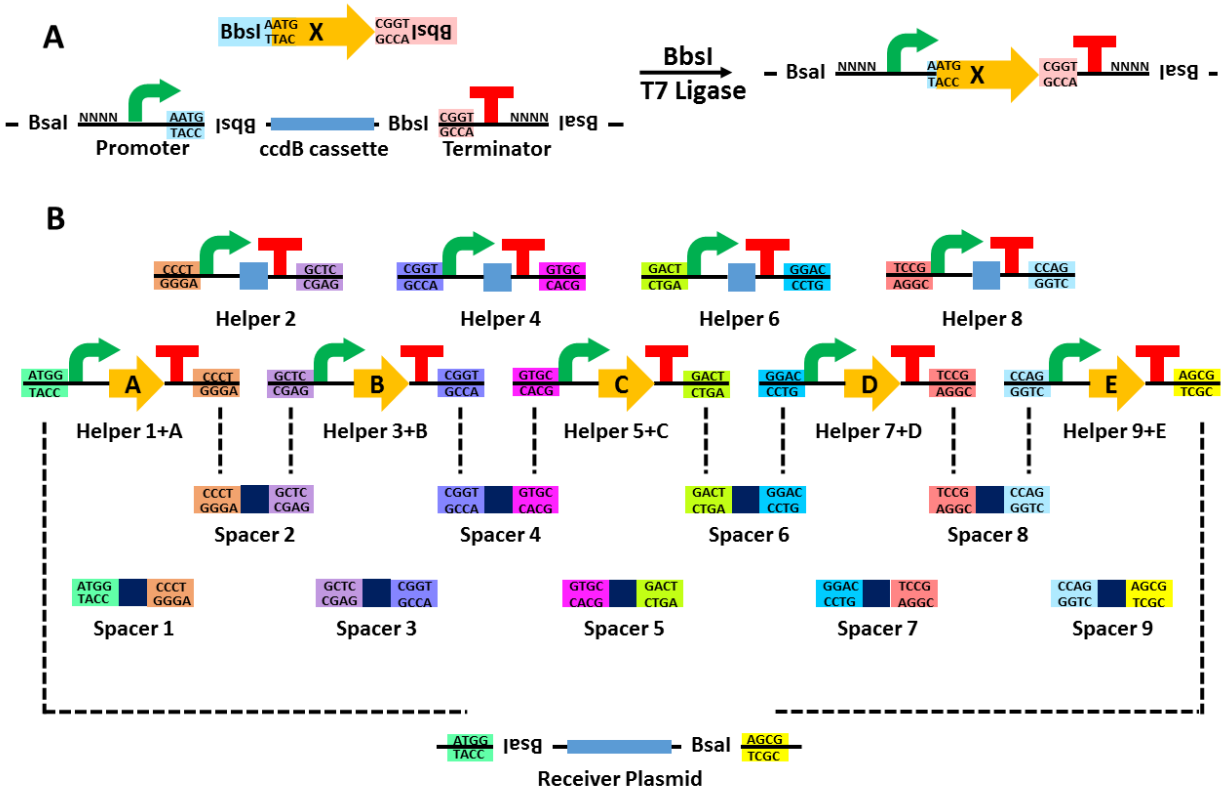


Figure 2.2. The fidelity of the 1st tier reaction checked by blue-white screening.

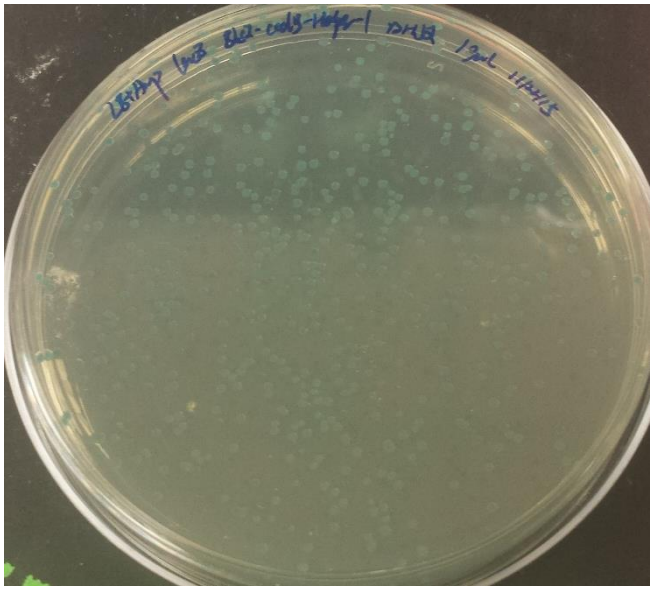


Figure 2.3. (A) The fidelity of the 1st tier reaction (monoclonal) checked by restriction digestion (expected band sizes: CrtE-Helper 1: 1.7 kb + 1.7 kb; CrtB-Helper 3: 2.2 kb + 1.7 kb; CrtI-Helper 5: 2.6 kb + 1.7 kb; CrtY-Helper 7: 1.9 kb+ 1.7 kb; CrtZ-Helper 9: 1.7 kb + 1.6 kb). (B) The fidelity of the 1st tier reaction (polyclonal) checked by restriction digestion. All the plasmids obtained from the 1st tier reactions were double-checked by Sanger sequencing. (C-E) The fidelity of the 2nd tier reaction (scenarios I-IV) checked by restriction digestion (expected band sizes: 5.7 kb + 4.3 kb + 2.7 kb + 1.6 kb). (F) The crude extract of zeaxanthin producing CEN.PK2-1C. (G) HPLC profile of zeaxanthin producing CEN.PK2-1C strains. This figure was adapted from (23).

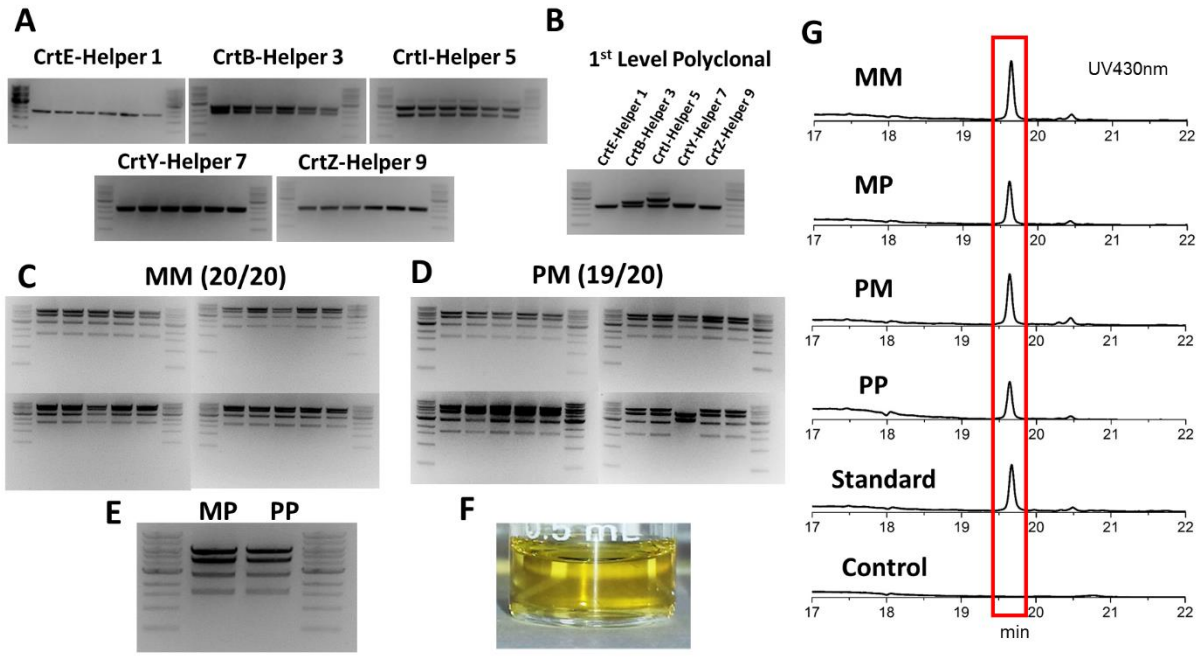


Figure 2.4. Analysis of the zeaxanthin producing pathway variants with gene deletion expressed in *S. cerevisiae*. (A) Pathways refactored for analysis. (B) Structures of carotenoids expected to be produced. (C) HPLC-UV traces. (D) UV absorption of the corresponding peaks. MS data is shown in Figure 2.6. This figure was adapted from (23).

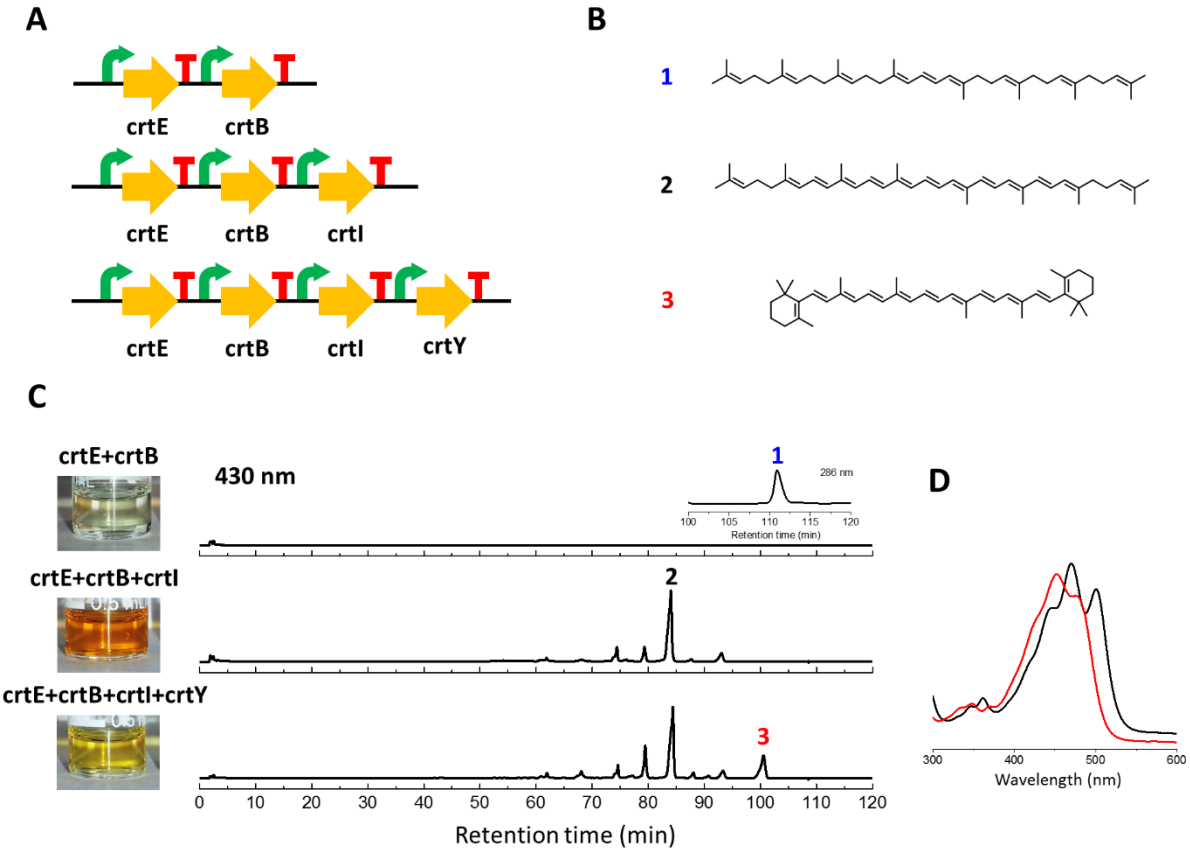


Figure 2.5. Analysis of the C30 carotenoid producing pathways expressed in *E. coli*. (A) Pathways refactored for analysis. (B) Structures of carotenoids expected to be produced. (C) HPLC-UV traces. (D) UV absorption of the corresponding peaks. MS data is shown in Figure 2.6. This figure was adapted from (23).

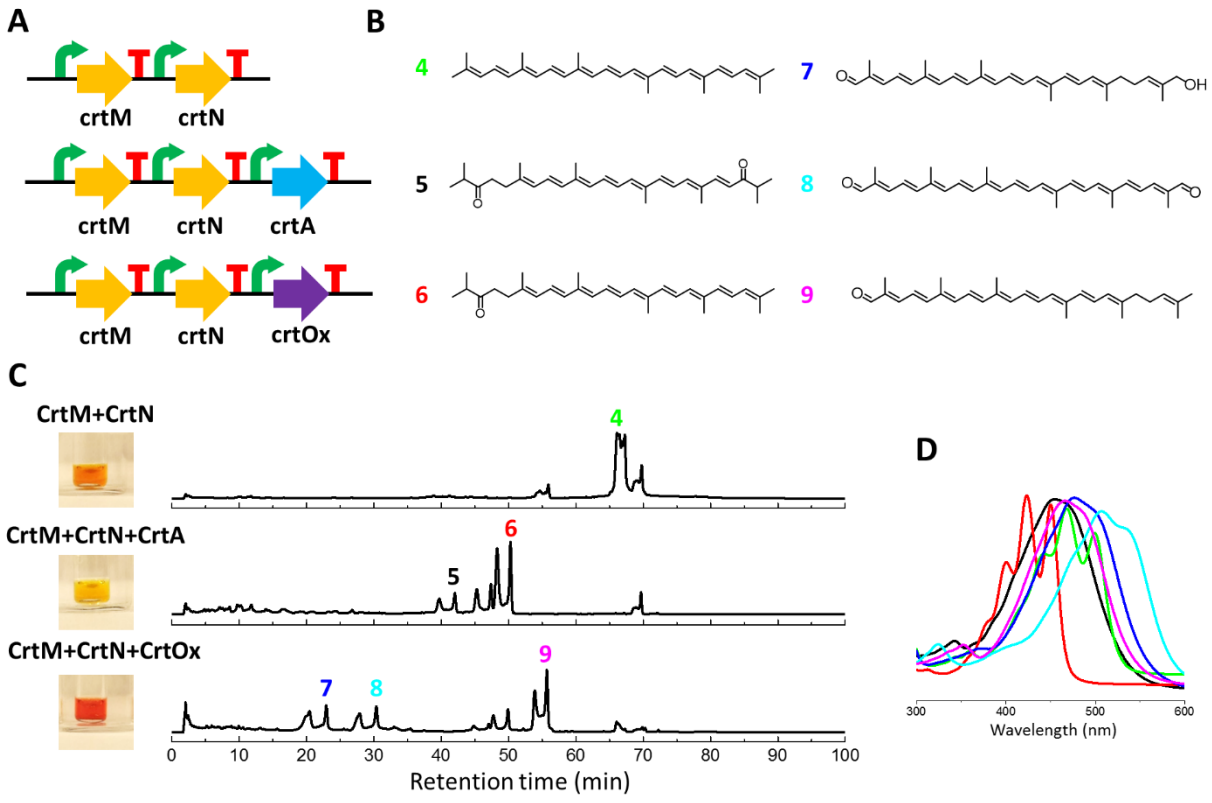
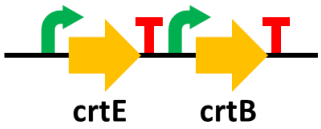
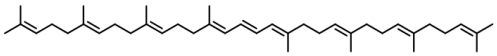
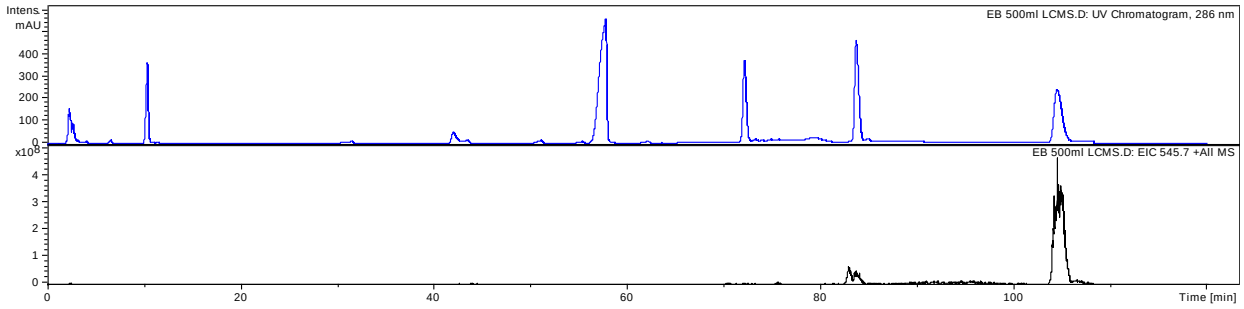


Figure 2.6. The HPLC profile and corresponding extracted ion chromatography obtained from liquid chromatography-mass spectrometry (LC/MS) for various carotenoid pathways.

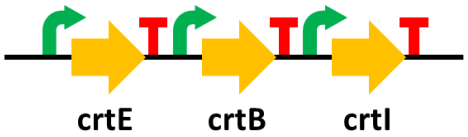
CrtE+CrtB

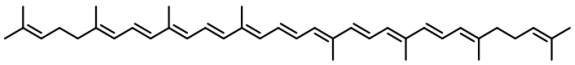


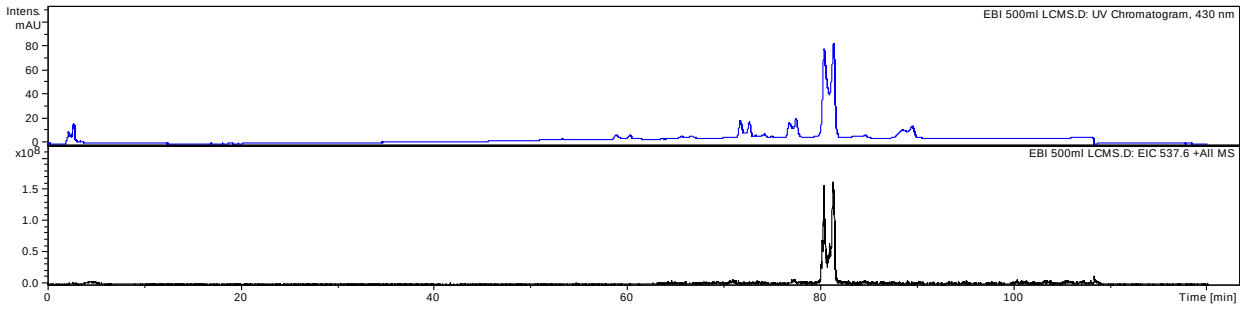
1 



CrtE+CrtB+CrtI



2 



CrtE+CrtB+CrtI+CrtY

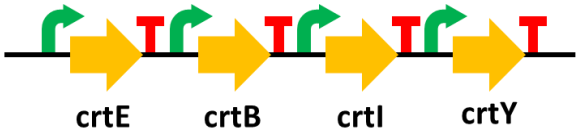
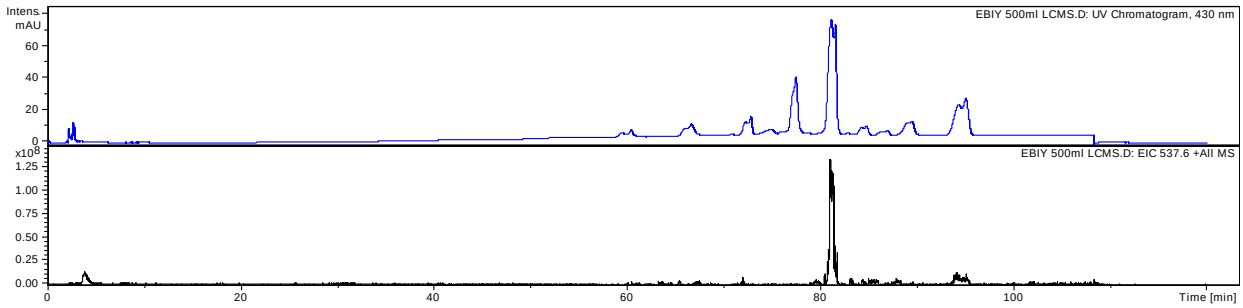
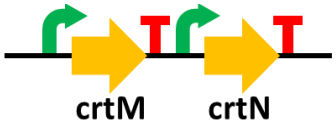


Figure 2.6. Continue

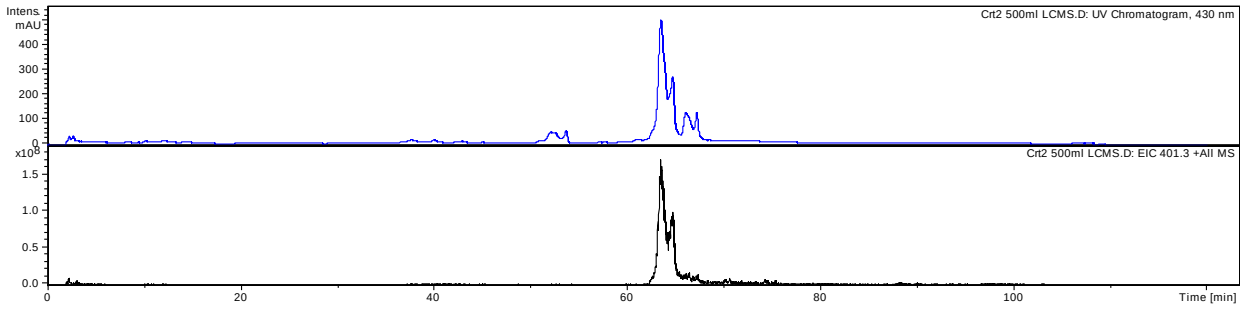
3



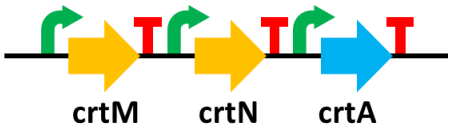
CrtM+CrtN



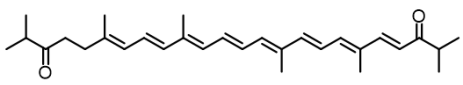
4



CrtM+CrtN+CrtA



5



6

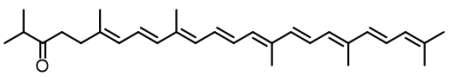
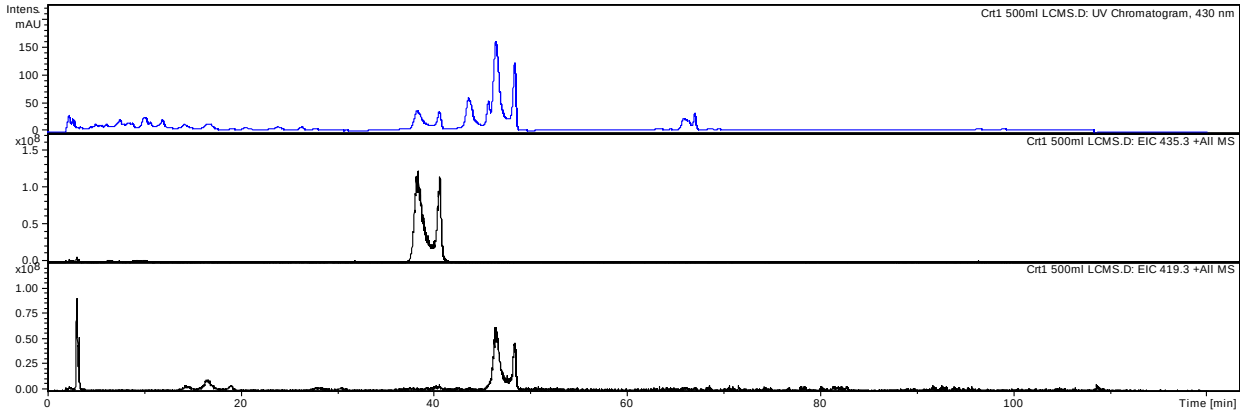
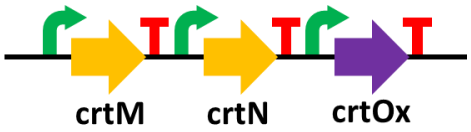


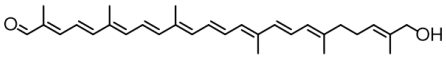
Figure 2.6. Continue



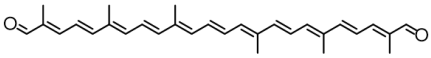
CrtM+CrtN+CrtOx



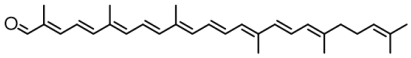
7



8



9



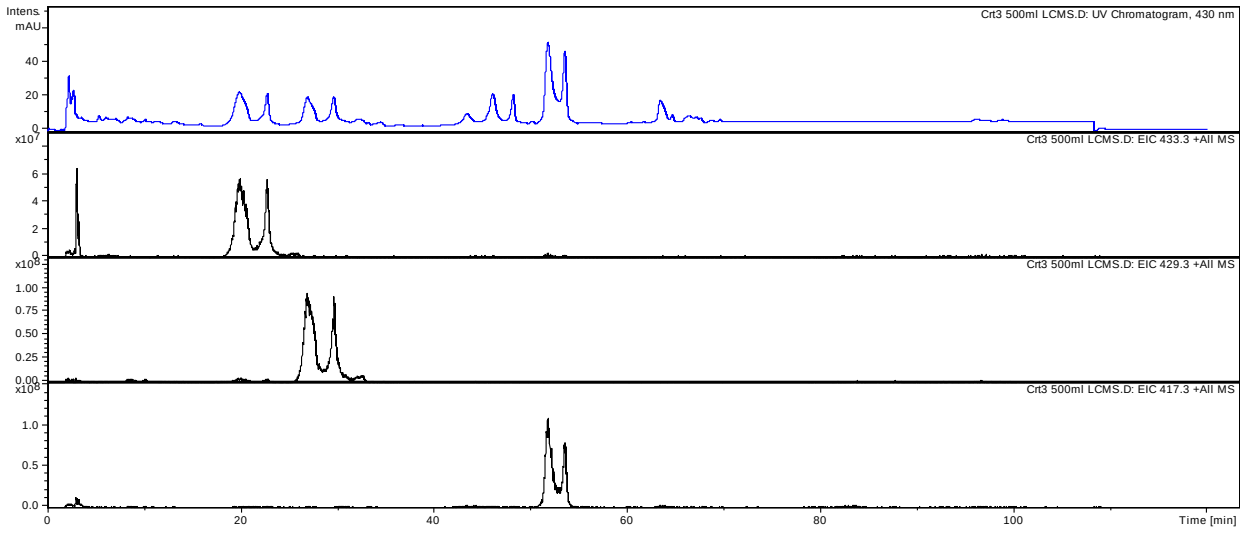


Figure 2.7. (A) The genes included in the 96 combinatorial pathways for carotenoid biosynthesis. Negative control (BL21(DE3) with empty pET28a plasmid) at the right side. (B) Cell pellets of BL21(DE3) after the induction by IPTG. This figure was adapted from (23).

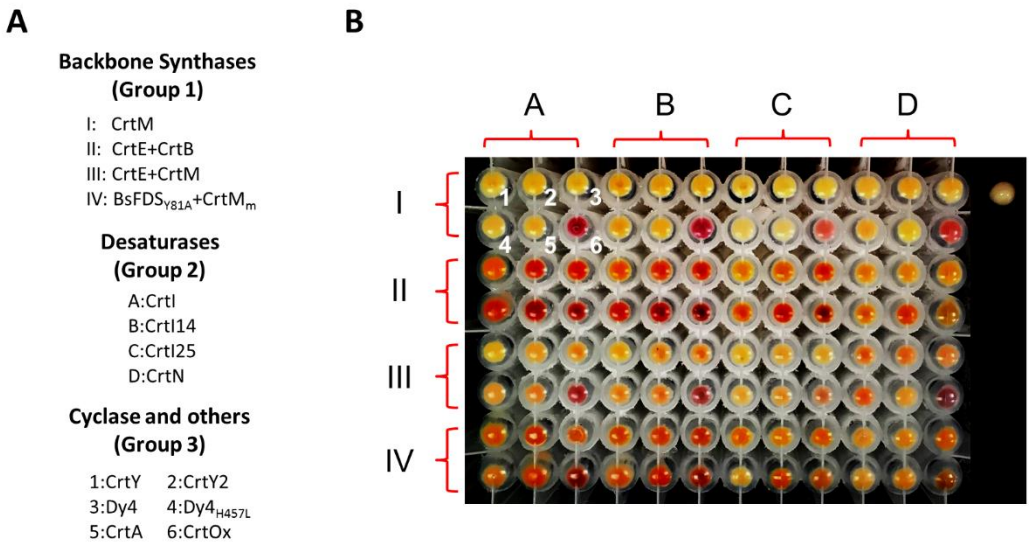
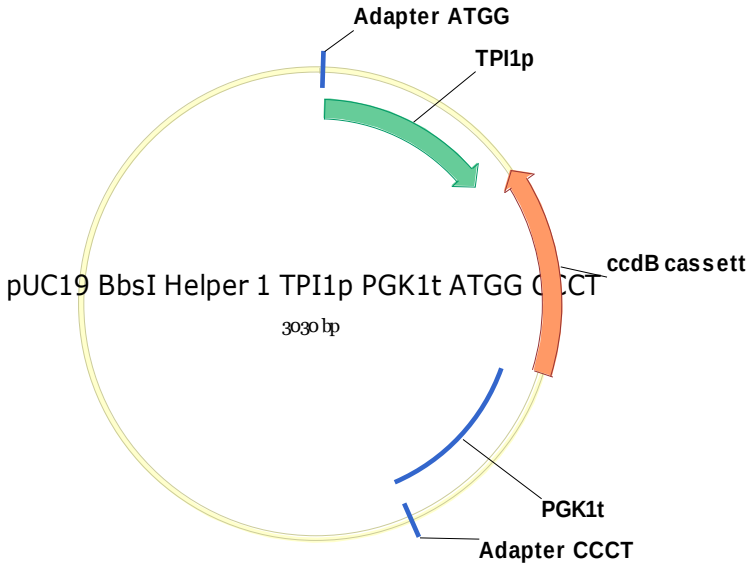


Figure 2.8. Plasmid designs and sequences of the *S. cerevisiae* helper plasmids.

pUC19 BbsI ccdB Helper 1 TPI1p PGK1t ATGG CCCT



pUC19 BbsI ccdB Helper 2 ENO2p GPDt CCCT GCTC

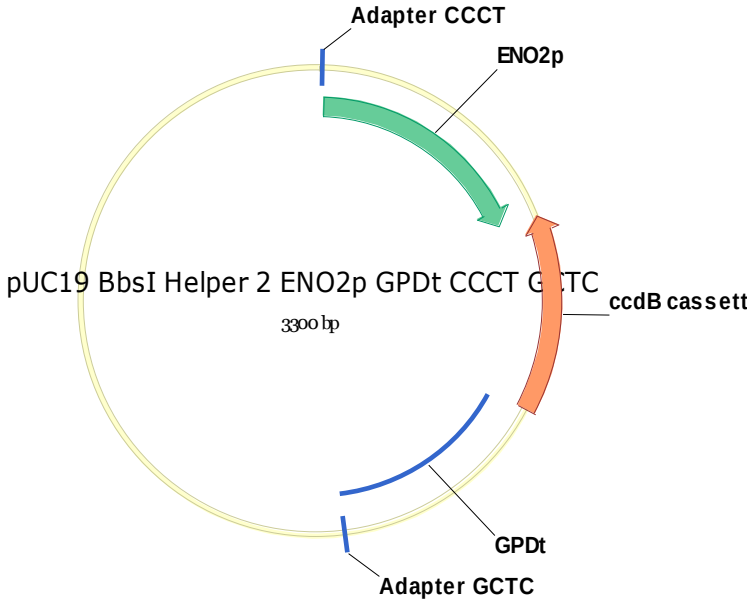
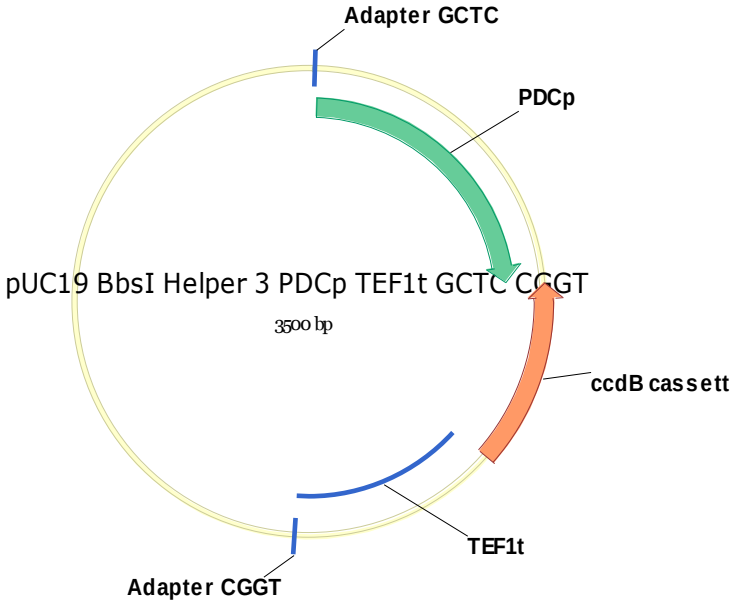


Figure 2.8 Continue

pUC19 BbsI ccdB Helper 3 PDCp TEF1t GCTC CGGT



pUC19 BbsI ccdB Helper 4 FBA1p HXT7t CGGT GTGC

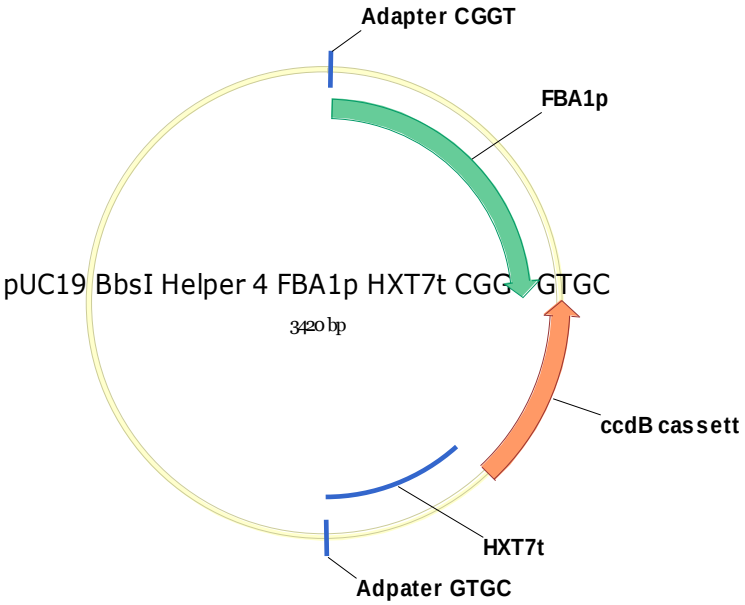
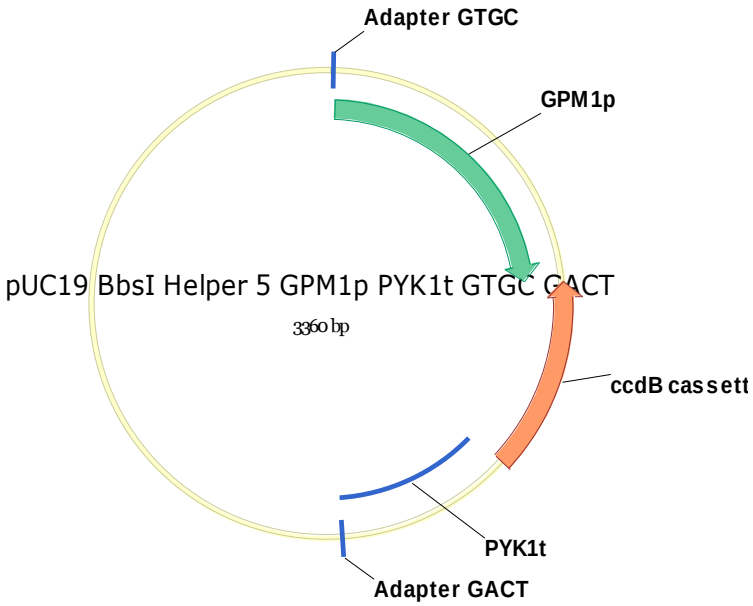


Figure 2.8 Continue

pUC19 BbsI ccdB Helper 5 GPM1p PYK1t GTGC GACT



pUC19 BbsI ccdB Helper 6 HXT7p ADH2t GACT GGAC

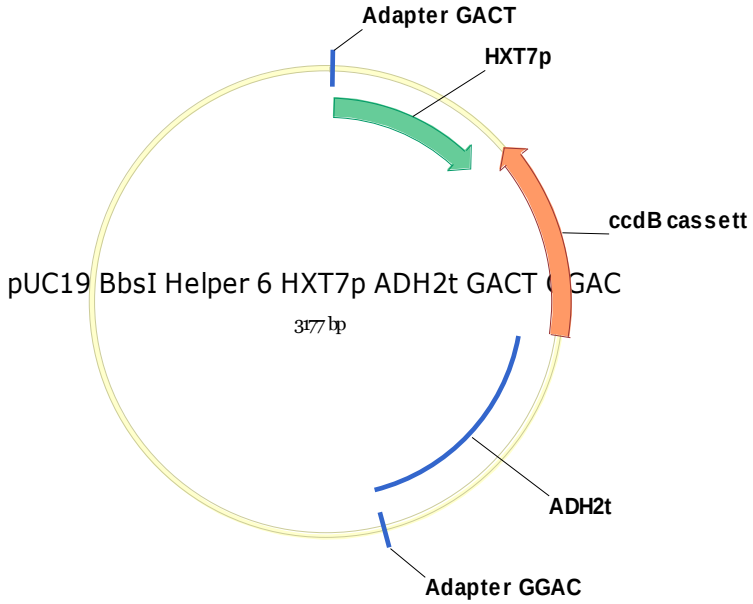
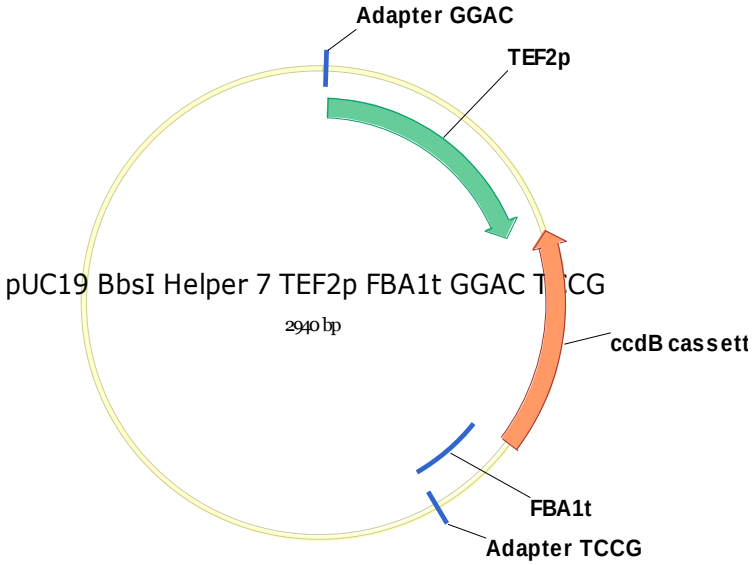


Figure 2.8 Continue

pUC19 BbsI ccdB Helper 7 TEF2p FBA1t GGAC TCCG



pUC19 BbsI ccdB Helper 8 TDH2p ADH1t TCCG CCAG

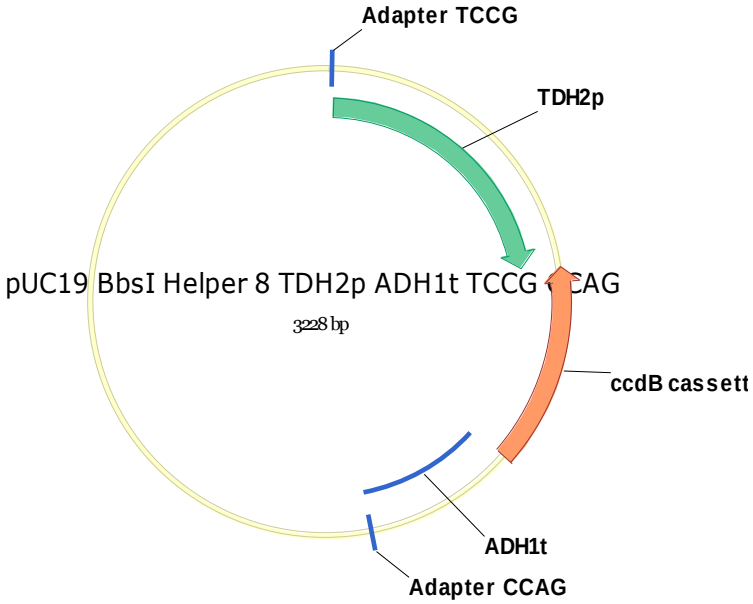


Figure 2.8 Continue

pUC19 BbsI ccdB Helper 9 GPDp GPM1t CCAG AGCG

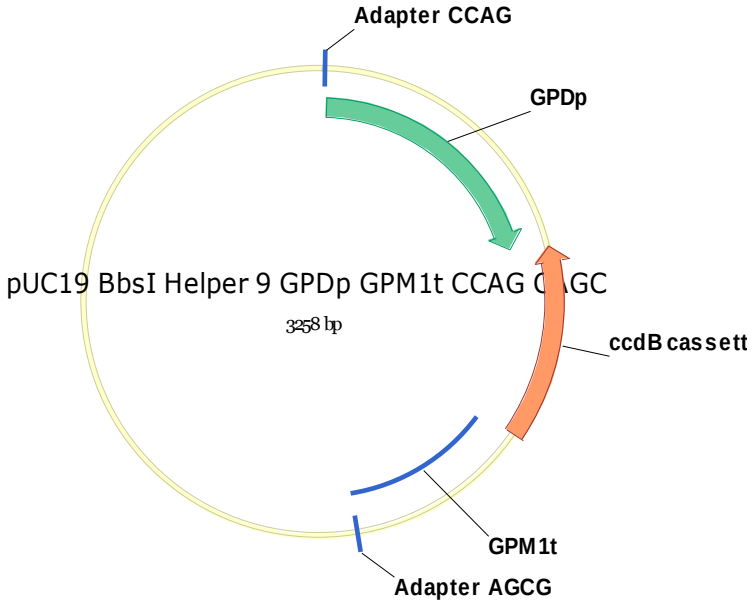


Figure 2.9. The plasmid design and sequence of the T7 helper plasmid-1. Other T7 helper plasmids share the same design and sequence except for the 4 bp overhangs, which are listed in Table 1.

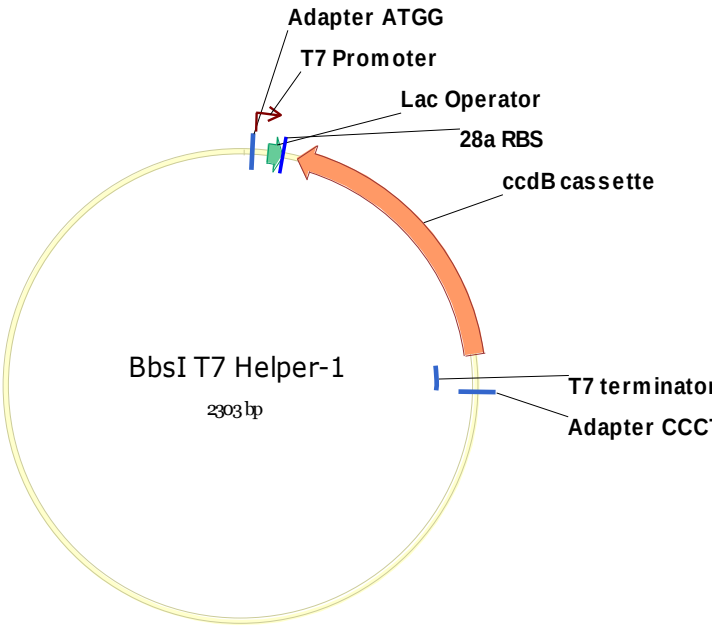


Figure 2.10. The plasmid design and sequence of the spacer-1. Other spacer plasmids share the same design and sequence except for the 20 bp spacer sequences and 4 bp overhangs, which are listed in Table 1.

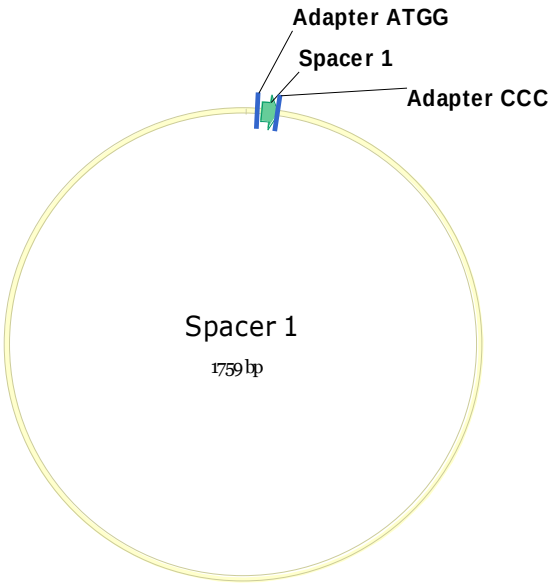


Figure 2.11. The plasmid design and sequence of the receiver plasmid for expression in *S. cerevisiae.*

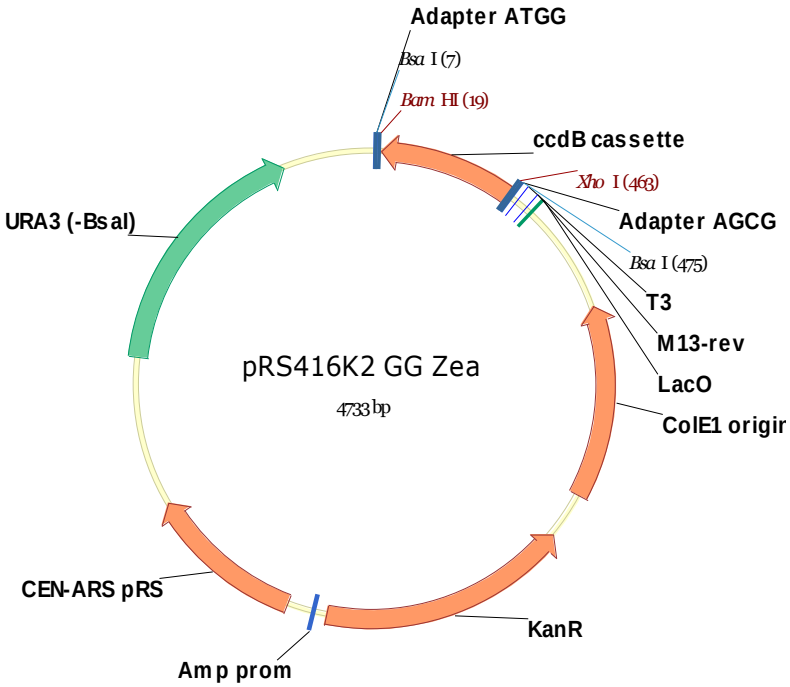


Figure 2.12. The plasmid design and sequence of the receiver plasmid for expression in *E. coli*

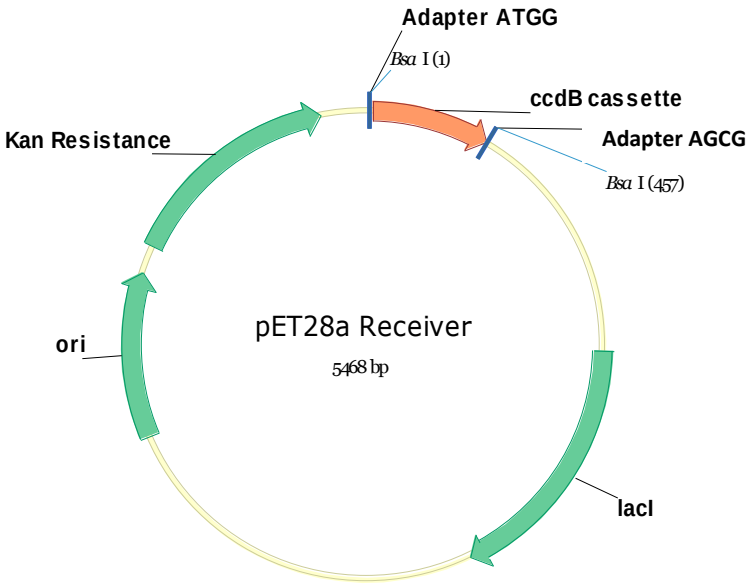


Table 2.1. Promoters and terminators used in construction of helper plasmids for expression in *S. cerevisiae* and the corresponding 20 bp spacer sequences with 4 bp overhangs. All of the helper plasmids for expression in *E. coli* use the T7 promoter and terminator and the 4th helper plasmid uses the CGGT and AGCG as the 4 bp overhangs.

Position	Promoter	Terminator	20 bp Spacer Sequence+4 bp Overhangs
1	TPI1P	PGK1t	ATGGccgagttcgctggtcacactCCCT
2	ENO2p	GPDt	CCCTagacgctgtgcctctcagtaGCTC
3	PDCp	TEF1t	GCTCctatcagcgtcactgggtccCGGT
4	FBA1p	HXT7t	CGGTggtttcacaggtccctccagGTGC
5	GPM1p	PYK1t	GTGCcctatttcaggacgctcggcGACT
6	HXT7p	ADH2t	GACTctcgcacctggctgagttacGGAC
7	TEF2p	FBA1t	GGACcacagggtcgtattgcctccTCCG
8	TDH2p	ADH1t	TCCGcgcctcgtgtgctatccagaCCAG
9	GPDp	GPM1t	CCAGgtcgcccgagtaacctctgtAGCG

2.6 References

1. Newman DJ & Cragg GM (2016) Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79(3):629-661.
2. Sun H, Liu Z, Zhao H, & Ang EL (2015) Recent advances in combinatorial biosynthesis for drug discovery. *Drug Des Devel Ther* 9:823-833.
3. Weber T & Kim HU (2016) The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotech* 1(2):69-79.
4. Gomez-Escribano JP, Alt S, & Bibb MJ (2016) Next generation sequencing of actinobacteria for the discovery of novel natural products. *Mar Drugs* 14(4):78.
5. Cobb RE, Ning JC, & Zhao H (2014) DNA assembly techniques for next-generation combinatorial biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.* 41(2):469-477.
6. Engler C, Gruetzner R, Kandzia R, & Marillonnet S (2009) Golden Gate shuffling: a one-pot DNA shuffling method based on Type IIs restriction enzymes. *Plos One* 4(5):e5553.
7. Liang J, Chao R, Abil Z, Bao ZH, & Zhao HM (2014) FairyTALE: a high-throughput TAL effector synthesis platform. *ACS Synth Biol* 3(2):67-73.
8. Shao Z, Zhao H, & Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37(2):e16.
9. Casini A, *et al.* (2014) R2oDNA designer: computational design of biologically neutral synthetic DNA sequences. *ACS Synth Biol* 3(8):525-528.
10. Schmidt-Dannert C, Umeno D, & Arnold FH (2000) Molecular breeding of carotenoid biosynthetic pathways. *Nat Biotechnol* 18(7):750-753.
11. Umeno D, Tobias AV, & Arnold FH (2005) Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiol Mol Biol Rev* 69(1):51-78.
12. Lee PC, Holtzapple E, & Schmidt-Dannert C (2010) Novel activity of Rhodobacter sphaeroides spheroidene monooxygenase CrtA expressed in *Escherichia coli*. *Appl Environ Microbiol* 76(21):7328-7331.
13. Mijts BN, Lee PC, & Schmidt-Dannert C (2005) Identification of a carotenoid oxygenase synthesizing acyclic xanthophylls: combinatorial biosynthesis and directed evolution. *Chem Biol* 12(4):453-460.
14. Schmidt-Dannert C, Lee PC, & Mijts BN (2006) Creating carotenoid diversity in *E. coli* cells using combinatorial and directed evolution strategies. *Phytochem Rev* 5(1):67-74.
15. Umeno D & Arnold FH (2003) A C35 carotenoid biosynthetic pathway. *Appl Environ Microbiol* 69(6):3573-3579.
16. Umeno D & Arnold FH (2004) Evolution of a pathway to novel long-chain carotenoids. *J. Bacteriol.* 186(5):1531-1536.
17. Cunningham FX, Jr. & Gantt E (2001) One ring or two? Determination of ring number in carotenoids by lycopene epsilon-cyclases. *Proc Natl Acad Sci U S A* 98(5):2905-2910.
18. Sarrion-Perdigones A, *et al.* (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *Plos One* 6(7):e21622.
19. Weber E, Engler C, Gruetzner R, Werner S, & Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. *Plos One* 6(2):e16765.

20. Iverson SV, Haddock TL, Beal J, & Densmore DM (2016) CIDAR MoClo: improved MoClo assembly standard and new E. coli part library enable rapid combinatorial design for synthetic and traditional biology. *ACS Synth Biol* 5(1):99-103.
21. Lee ME, DeLoache WC, Cervantes B, & Dueber JE (2015) A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol* 4(9):975-986.
22. Luo Y, *et al.* (2013) Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nat Commun* 4:2894.
23. Ren H, Hu P, & Zhao H (2017) A plug-and-play pathway refactoring workflow for natural product research in Escherichia coli and Saccharomyces cerevisiae. Biotechnol. Bioeng. 114(8):1847-1854..

Chapter 3: Rapid Discovery and Characterization of Glycocins Through Pathway Refactoring in *E. coli*

3.1 Introduction

Ribosomally synthesized and post-translationally modified peptides (RiPPs) constitute an emerging family of natural products with various structural features and bioactivities (1). Glycocins (glycosylated bacteriocins) are a group of RiPPs that feature sugar moieties installed on serine, threonine, or cysteine residues and two nested disulfide bonds that stabilize their helix-loop-helix structures. A typical biosynthetic pathway for glycocin involves five genes. The gene encoding the precursor peptide is expressed first and the resulting precursor peptide is then glycosylated by a glycosyltransferase. Two disulfide oxidoreductases are responsible for forming the disulfide bonds, and a bifunctional transporter transfers the peptide outside the cell and removes the leader peptide (2). Four groups of glycocins with distinct scaffolds have been discovered to date, with the prototypes being glycocin F (GccF-type), sublancin (SunA/ThuA type), enterocin 96 (enterocin 96-type), and enterocin F4-9 (enterocin F4-9-type) (2).

Owing to recent advances in next-generation sequencing and bioinformatics, many putative glycocin biosynthetic gene clusters (BGCs) have been identified through genome mining. However, only six glycocins have been discovered so far: five isolated from native producers and one by chemoenzymatic synthesis (2). This dearth of characterized glycocins may be attributed to our limited knowledge of glycocins and technical challenges in product isolation from the native producers, particularly pathogens and extremophiles. All six known glycocins show antimicrobial bioactivities against several Gram-positive pathogens such as *Bacillus cereus*, *Streptococcus pyogenes*, and *Staphylococcus aureus*. Particularly, one of them, sublancin, was also active against

drug resistant strains such as methicillin–resistant *S. aureus* (MRSA) and gentamicin–resistant *Enterococcus faecalis* (3-5). Therefore, glycocins appear to be a promising source for drug lead discovery (3, 6) and it is desirable to discover novel glycocins.

In this work, we sought to use a synthetic biology approach based on the pathway refactoring strategy we previously developed (7) to rapidly discover new glycocins. RODEO (Rapid ORF Description and Evaluation Online) (8) a powerful genome mining tool for RiPPs, was used for predicting putative glycocin BGCs. According to their distinct features, six of the putative glycocin BGCs were chosen for pathway refactoring in *Escherichia coli*. Four of them successfully produced new glycocins, which were named bacillicin CER074, bacillicin BAG2O, geocillicin, and listeriocytocin, respectively. Although these glycocin BGCs originated from various species and most of them are either pathogens or extremophiles, our synthetic biology strategy involving DNA synthesis and pathway refactoring in *E. coli* has made the glycocin discovery process rapid and scalable.

3.2 Results and Discussion

3.2.1 Design of a Modified Pathway Refactoring Workflow

Previously we developed a plug-and-play pathway refactoring workflow for rapid discovery of natural products (7). This workflow consists of two-step Golden Gate reactions that first convert biosynthetic genes into individual expression cassettes and then assemble these expression cassettes into one plasmid. In this work, we adapted this workflow for discovery of new glycocins (Figure 3.1). Briefly, the target glycocin BGC is first partially refactored, which involves only the peptide precursor gene and the glycosyltransferase gene that are required to generate a

glycosylated intermediate (first tier) (9). If a compound corresponding to the expected intermediate is detected, the corresponding BGC will then be fully refactored to generate the final product (second tier). *E. coli* BL21(DE3) is used as a heterologous expression host and the BGC genes are codon optimized for *E. coli* and obtained through DNA synthesis. For the refactored partial glycocin BGC, after overnight culture in 1 mL media under various temperatures, the resulting cell pellets are obtained by centrifugation and homogenized in an ultrasonic bath. For the refactored full glycocin BGC, supernatants instead of cell pellets were used for next step analysis. Both the supernatants and the cell lysates are desalted and analyzed by MALDI-MS (Matrix Assisted Laser Desorption/Ionization-Mass Spectrometry). While this workflow is designed for the discovery of glycocins, it's also potentially applicable for high-throughput discovery of other RiPP families in the future.

3.2.2 Refactoring of Glycocin BGCs for Heterologous Expression in *E. coli*

Putative glycocin BGCs were predicted by RODEO. The amino acid sequence of glycosyltransferase for sublancin (SunS) was used as the query for BLAST and the top 5000 hits were used as the input for RODEO, which analyzed the neighboring genes of each input gene in batch through comparison against the Pfam pHMM database. All putative glycosyltransferases with their neighboring genes were then filtered by the typical features of the glycocin BGCs as described in the Materials and Methods. A sequence similarity network was then generated for the putative precursor peptides and six BGC candidates for refactoring were mainly chosen from different groups in this network (Figure 3.2A). Two out of the six glycocin BGCs which according to the sequence of the precursor peptide belong to the same group as the BGC from *Bacillus megaterium* BHG1.1, were chosen because of their extremophile origin (thermophilic strain,

Geobacillus sp. 8 and alkaliphilic strain, *Bacillus* sp. JCM19047), which may offer better stabilities under certain conditions for their future studies and/or applications (Figure 3.2 and Figure 3.3) (10-14).

The first tier pathway refactoring was performed as indicated in Figure 3.1B. Both the precursor peptide and glycosyltransferase were expressed as individual expression cassettes with T7 promoter and T7 terminator. After the first tier pathway refactoring process, three out of the six partial glycocin BGCs expressed in *E. coli* generated peaks in the MALDI-TOF mass spectrum with molecular weights corresponding to the monoglycosylated precursor peptides (Figure 3.4A). Based on the mass increase (162 Da), all the installed sugars were hexoses. The partial glycocin BGC from *Listeria monocytogenes* generated a new peak in the mass spectrum consistent with a diglycosylated precursor peptide with an N-terminal truncation. To confirm this result, we introduced site-specific mutations (K7A and M8A) at the putative cleavage sites on its precursor peptide and coexpressed the mutant precursor gene with the glycosyltransferase (Figure 3.2B). The expected full length mutant precursor peptide with two hexoses installed was indeed observed in the MALDI-TOF mass spectrum (Figure 3.4B). Given the successful glycosylations, all four glycocin BGCs were then fully refactored with every biosynthetic gene expressed as individual expression cassettes under T7 promoter and T7 terminator (Figure 3.1B). Of note, for the glycocin BGC from *Listeria monocytogenes*, the mutant precursor peptide gene was used in pathway refactoring. New peaks in the mass spectrum corresponding to the glycosylated precursor peptides with their leader peptides removed at the expected sites (Figures 3.3B and 3.6) were observed from the desalted supernatants. The resulting four glycocins were named bacillicin CER074, bacillicin BAG2O, geocillicin, and listeriocytocin. According to the classification by Norris and coworkers

(2), bacillicin CER074, bacillicin BAG2O and geocillicin are SunA/ThuA-type glycocins (10, 13), while listeriocytocin is a member of the enterocin 96-type glycocins (Figure 3.2C) (12).

3.2.3 Isolation and Structural Characterization of the Glycocins

In order to obtain sufficient amounts of glycocins for structure characterization and bioassays (at least 40 μg for each glycocin for both MS based structural characterization and standard agar diffusion assay for bioactivity screening), we further optimized cultivation conditions and purified the glycocins through ammonium sulfate precipitation, solid phase extraction, semiprep HPLC, and analytical HPLC (Figure 3.7B). The existence of disulfide bonds was first confirmed by an iodoacetamide (IAA) assay that demonstrated the absence of free thiols (Figure 3.8) (9). The purified glycocins were then digested by chymotrypsin under non-reducing and reducing conditions and analyzed by MS. Through assignment of the observed fragments, we concluded that all four glycocins shared similar topologies (Figure 3.9). Although the glycosylated residues for the three SunA/ThuA type glycocins can be easily assigned to the third cysteine by the results of iodoacetamide derivatization and proteolytic digest, MS/MS analysis was necessary to determine the glycosylated residue in listeriocytocin. Through sequence alignment, Ser60 in the precursor peptide of listeriocytocin was predicted to be glycosylated (Figure 3.3B). MS/MS data unambiguously confirmed this prediction and showed that two successive glycosylations occur on Ser60 (Figure 3.10, Figure 3.11, and Figure 3.12).

To further determine the type of hexoses installed in the four glycocins, the glycosylated precursor peptides were purified from the corresponding partial BGCs, which resulted in a higher yield than refactoring of the full BGCs (Figure 3.5). Notably, in the first tier pathway refactoring, the

precursor genes were cloned into a helper plasmid with an internal N-terminal His-Tag at the cloning site (Figure 3.13), which enables the purification of the glycosylated intermediates by IMAC (immobilized metal affinity chromatography). These intermediates were then further purified by RP-HPLC (reversed-phase high-performance liquid chromatography) and hydrolyzed. The sugars were derivatized and analyzed using gas chromatography as described elsewhere (9, 15). In comparison with various hexose standards, all the installed sugars on these four glycocins were shown to be glucoses (Figure 3.7D). Previously, two glycosyltransferases for glycocin biosynthesis have been characterized in vitro and exhibited promiscuous activity towards different nucleotide sugar substrates (15, 16). Considering that the concentrations of some potential nucleotide sugar substrates, such as UDP-N-acetylglucosamine, are much higher than UDP-glucose in *E. coli* (17, 18), it appears that the glycosyltransferases under investigation here have a preference for UDP-glucose. To further investigate if glycosyltransferases for glycocin biosynthesis may exhibit high specificity for their nucleotide sugar substrates *in vivo*, we also coexpressed the precursor peptide of enterocin F4-9 with its glycosyltransferase in BL21(DE3), and mass which represents the peptide modified with two N-acetylglucosamines was clearly observed on MALDI-MS (Figure 3.14).

All three SunA/ThuA-type glycocins shared similar topology and sugar modification as sublancin (Figure 3.9), suggesting these structural features are conserved in this glycocin family. Although enterocin 96 was the first discovered enterocin 96-type glycocin, its structure has yet to be elucidated (19). Recently the glycosyltransferase of enterocin 96 was reconstituted *in vitro* and the modification was shown to be a disaccharide (16). As we discovered that the listeriocytocin is also diglucosylated at the same serine residue predicted to be glycosylated through alignment to

enterocin 96, it is very likely that the wild type enterocin 96 shares a similar structure as that produced *in vitro* and the type of sugar modification might be glucose. The diglucosylation might be a distinct feature of enterocin 96-type glycocins from other diglycosylated glycocins, such as enterocin F4-9 and glycocin F (20, 21). However, listeriocytocin has very distinct structural features compared to enterocin 96, such as a much smaller loop size (Figure 3.2C).

3.2.4 Antimicrobial Activity of the Glycocins

The antimicrobial activity of the four glycocins was probed against a panel of Gram-positive and Gram-negative bacteria (*Escherichia coli* JM109, *Pseudomonas aeruginosa* PA01, Methicillin-resistant *Streptococcus aureus* NRS384/USA300, Vancomycin-resistant *Enterococcus faecium* V583 CB807, *Listeria monocytogenes* 4b F2365, *Bacillus subtilis* Δspβ, *Bacillus halodurans* C-125 and *Bacillus cereus* ATCC14579) by a standard agar diffusion assay (Table 1). Although listeriocytocin lacked antimicrobial activity toward all organisms tested, bacillicin CER074 and bacillicin BAG2O exhibited antimicrobial activity against *Bacillus cereus* ATCC 14579 (minimal inhibitory concentrations: sublancin, 2.5 μM; bacillicin CER074, 0.156 μM; bacillicin BAG2O, 0.625 μM), and geocillicin also exhibited weak bioactivity (Figure 3.15). Overall, the three newly discovered SunA/ThuA-type glycocins have narrower antimicrobial spectrum than sublancin with all of them inhibiting the growth of *B. cereus* ATCC14579, but not other bacilli tested (Figure 3.16 and Figure 3.17).

As the bioactivity of glycocin F, which is a glycocin decorated with two *N*-acetylglucosamine (GlcNAc) moieties, can be reversed by free GlcNAc (22), we tested if free glucose can have a similar effect on the antimicrobial activities of the four new glycocins. While the bioactivities of

sublancin and geocillicin could be reversed by free glucose, no significant effects were observed for bacillicin CER074 and bacillicin BAG2O, which means these two glycocins might have different target(s) from sublancin and geocillicin (Figure 3.5). Previous studies showed that sublancin's activity is mediated by the phosphoenolpyruvate:sugar phosphotransferase system (PTS) and the large mechanosensitive channel of conductance MscL (23, 24). In structure-activity relationship (SAR) studies, Arg33 of sublancin was previously shown to be critical for its bioactivity against *B. subtilis* Δspβ and most of the SunA/ThuA-type glycocin are positively charged at the corresponding site in a sequence alignment to sublancin (9). However, bioactivity was also observed for bacillicin BAG2O that does not contain a corresponding charged residue, which may imply that bacillicin BAG2O has a different target(s) from sublancin. Therefore, it would be of great interest to further investigate the antimicrobial mechanisms of these new glycocins, and our production platform in *E. coli* provides the opportunity to perform further SAR studies.

3.3 Conclusion

In this work, we used a rapid pathway refactoring workflow to discover four novel glycocins from various origins, which has significantly expanded the glycocin family. This workflow is potentially scalable, which can be very helpful for RiPP discovery in the future. We provide evidence that di-glycosylation might be a general feature for enterocin 96-type glycocins, and the bioactivity assays demonstrate that factors besides the topology and type of glycosylation may affect the bioactivities of glycocins, which is of great interest for further investigation of structure-function relationships.

3.4 Materials and Methods

3.4.1 Materials and Reagents

All chemicals were purchased from Sigma-Aldrich and Thermo Fisher Scientific. *E. coli* NEB® 10-beta was used for DNA manipulation and *E. coli* BL21(DE3) was used as a host for expression. Genes were obtained as gBlocks® from Integrated DNA Technologies. *Bsa*I, T4 DNA ligase and other restriction enzymes for checking the assembly of plasmids were obtained from New England Biolabs, while the *Bpi*I was obtained from Thermo Fisher Scientific and the T7 DNA ligase was obtained from Enzymatics. Plasmids were isolated using QIAGEN Plasmid Mini Kit. Kanamycin was used at 50 µg/ml for all *E. coli* cultures. ZipTip® C18 was purchased from EMD Millipore. All sequencing grade peptidases were obtained from Promega.

3.4.2 Genome Mining by RODEO

The amino acid sequence of SunS (the glycosyltransferase for the biosynthesis of sublancin) was used as the query for BLASTP, and the first 5000 hits were used as the input for the analysis by RODEO. Each predicted gene cluster was filtered by the existence of all genes which are necessary for glycocin biosynthesis (genes encoding SunA-like precursor, SunS-like glycosyltransferase, disulfide bond formation protein/thioredoxin, SunT-like transporter). The gene encoding a peptide less than 100 aa with at least four cysteines was considered as the putative precursor gene.

3.4.3 Pathway Refactoring Through the Plug-and-Play Workflow

The pathway refactoring process followed the protocol we described before with slight modifications (7). In the partial pathway refactoring, a new helper plasmid (T7 His Helper 1) with

an internal N-terminal His-tag at the cloning site was used in the first tier Golden Gate reaction for cloning the precursor gene, while the glycosyltransferase gene was cloned into the T7 Helper 2/9 plasmid. In the full pathway refactoring, the precursor gene, the glycosyltransferase gene, the disulfide bond formation protein/thioredoxin and the transporter gene were cloned into T7 Helper 1, T7 Helper 2, T7 Helper 3, T7 Helper 4, and T7 Helper 5/9 respectively. For the listeriocytocin BGC, two transporter genes (Figure 3.2) were cloned into T7 Helper 5 and T7 6/9, respectively.

3.4.4 Detection of the Glycosylated Precursor Peptide from Cell Pellets and Glycocin from Liquid Medium by MALDI-TOF MS

For the partial refactored BGCs, cells were grown in LB medium and IPTG was added to 0.5 mM when the OD600 reached 0.6-0.8. Cells were grown with shaking at 18 ºC, 30 ºC and 37 ºC in parallel for overnight induction. Cell pellets were harvested from 1 mL of liquid culture by centrifugation and resuspended in 100 µL of cell lysis buffer (8 M urea, 50 mM Tris • HCl, pH 7.5), which were then vigorously vortexed three times (30 s each time) and sonicated in ultrasonic bath (Branson 2800, Danbury, CT) for at least 30 min. The supernatants were then collected by centrifuge, desalted by ziptip and analyzed using MALDI-TOF MS. For the fully refactored BGCs, the supernatants of liquid culture were collected by centrifugation after overnight induction, desalted by ziptip and analyzed using MALDI-TOF MS. As rich media are too complex, which largely affect the product detection from the supernatant possibly due to ion suppression (data not shown), minimal medium M9 was used in the full BGC refactoring.

3.4.5 Preparation of the Glycocins

The *E. coli* BL21(DE3) transformant with the corresponding plasmid was streaked on an LB plate containing 50 μg/mL kanamycin and grew at 37 ºC overnight. A single colony was used to inoculate 50 mL fresh LB liquid medium and grew at 37 ºC 250 rpm for 12-14 h, which was used to inoculate fresh MOPS medium at 1:100 ratio for fermentation. The MOPS medium was prepared as described elsewhere (25). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to 0.5 mM when OD600 reached 0.6-0.8. For bacillicin CER074 and listeriocytocin, the temperature was then decreased to 18 ºC and maintained for 72 h. For bacillicin BAG2O, the temperature was decreased to 30 ºC and maintained for 144 h. For geocillicin, the temperature was maintained at 37 ºC for 144 h. The pH of the liquid culture was then adjusted to 2.5 by phosphonic acid and the supernatant was collected by centrifuging at 10000 × g for 15 min. The supernatant was cooled down to 4 ºC and ammonium sulfate was added slowly with stirring to reach 75% saturation. The precipitant was collected after overnight incubation under 4 ºC by centrifuging at 10000 × g for 15 min and resuspended in 0.1% trifluoroacetic acid (TFA). Insoluble materials were removed by centrifuging at 4000 × g for 20 min and loaded to BondElute C18 solid-phase extraction column (Agilent Technologies, Santa Clara, CA) previously wetted with 100% methanol and equilibrated by 5% B (solvent A = 0.1% TFA in HPLC grade water; solvent B = 0.1% TFA in HPLC grade acetonitrile/water). The glycocin was eluted using a step gradient with increasing percentage of solvent B in 50 mL volumes: 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% and 100% B. Fractions containing glycocin were combined and lyophilized to dryness.

Semi preparative HPLC was performed on an Agilent 1000 series HPLC system equipped with a Phenomenex C5 column (5 μm, 100 Å, 250 mm × 10 mm). The glycocin sample from the solid

phase extraction was dissolved in 0.1% TFA and applied to the column. The glycocin was eluted by a linear gradient from 5% B to 60 % B over 25 min with a flow rate of 4.0 mL/min. The glycocin containing peak was collected and lyophilized again, which was further purified by a Phenomenex C18 column (5 μm, 100 Å, 250 mm × 4.6 mm) by a linear gradient from 20% B to 40% B over 40 min.

Multiple mediums and induction time were used to optimize the culturing condition and MOPS with 3 (bacillicin CER074, listeriocytocin) or 6 days (bacillicin BAG2O, geocillicin) induction turned out to give the highest productivity (10-20 μg/L for bacillicin CER074, listeriocytocin and bacillicin BAG2O, 1 μg/L for geocillicin).

3.4.6 Preparation of the Glycosylated Intermediates

E. coli BL21(DE3) with the corresponding partial BGC was streaked on an LB plate with kanamycin and grew at 37 ºC overnight. A single colony was used to inoculate 50 mL fresh LB liquid medium and grew at 37 ºC 250 rpm for 12-14 h as the seed culture, which was used to inoculate 4L LB medium. The cell culture was induced when OD600 reached 0.6-0.8 by 0.5 mM IPTG. After 18 h induction under different temperatures (18 ºC for the bacillicin CER074 intermediate and mynocytogin intermediate, 30 ºC for bacillicin BAG2O intermediate and 37 ºC for geocillicin intermediate), the cell pellet was harvested by centrifuging at 10000 × g for 15 min. The intermediate was then prepared by IMAC (immobilized metal affinity chromatography) as described previously (26). The eluted fraction was desalted by BondElute C18 solid phase extraction column and lyophilized to dryness, which was then purified by a Phenomenex C5 HPLC

column (5 μm, 100 Å, 250 mm × 10 mm) by a linear gradient from 5% B to 60 % B over 25 min with a flow rate of 4.0 mL/min, as described previously.

3.4.7 Structure and Bioactivity Characterizations of Glycocins

IAA assays were performed as described previously with slight modifications (9). Briefly, reactions containing TCEP (tris(2-carboxyethyl)phosphine) were heated for 2 min at 55 °C, cooled to room temperature and then iodoacetamide was added. While the chymotrypsin digestion on Glycocins, determination of the sugar modification, MS/MS analysis and bioassays were performed as described elsewhere (9).

3.5 Figures and Tables

Figure 3.1. The workflow for rapid discovery of glycocins. A) The glycocin BGCs were predicted from genomic sequences that are available in NCBI (National Center for Biotechnology Information) by RODEO, and genes for the glycocin biosynthesis were annotated. B) The candidate BGCs were then refactored using a two-tier strategy and the resulting products were detected using MALDI-TOF MS. This figure was adapted from (27).

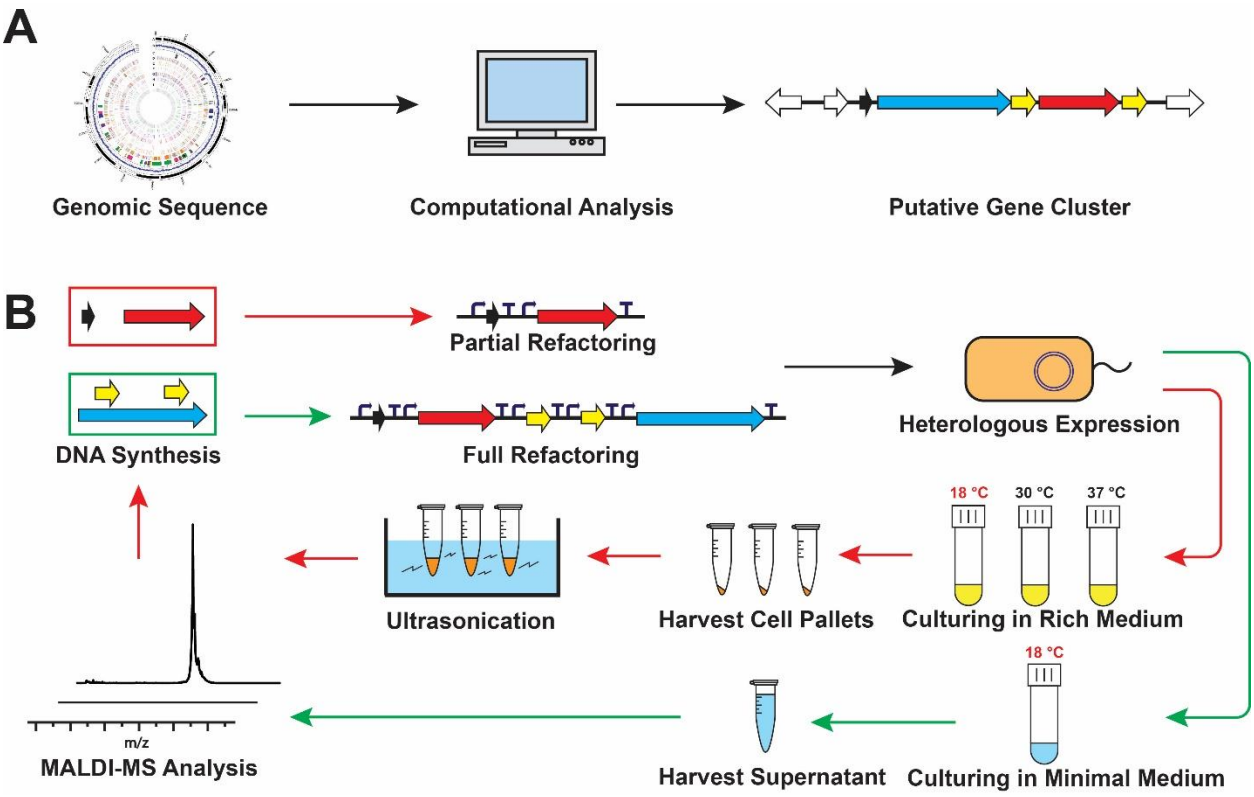


Figure 3.2. A) Sequence Similarity Network (SSN) of putative glycocin precursor peptides. Network was visualized with an edge cutoff E-value threshold of 10^{-8}, with nodes colored by genus. Location within the network of known glycocins (including those from this study) are indicated. B) Precursor peptide sequences and (proposed) structures for glycocin F, sublancin, thurandacin, enterocin 96 and enterocin F4-9. The cysteines for disulfide bond formation are highlighted in yellow and glycosylated residues are highlighted in purple. The boundary of leader peptide and core peptide is indicated by a red arrow. C) Alignment of the glycocins discovered in this work as different types with the corresponding known glycocins. The cysteines for disulfide bond formation are highlighted in yellow and glycosylated residues are highlighted in purple.

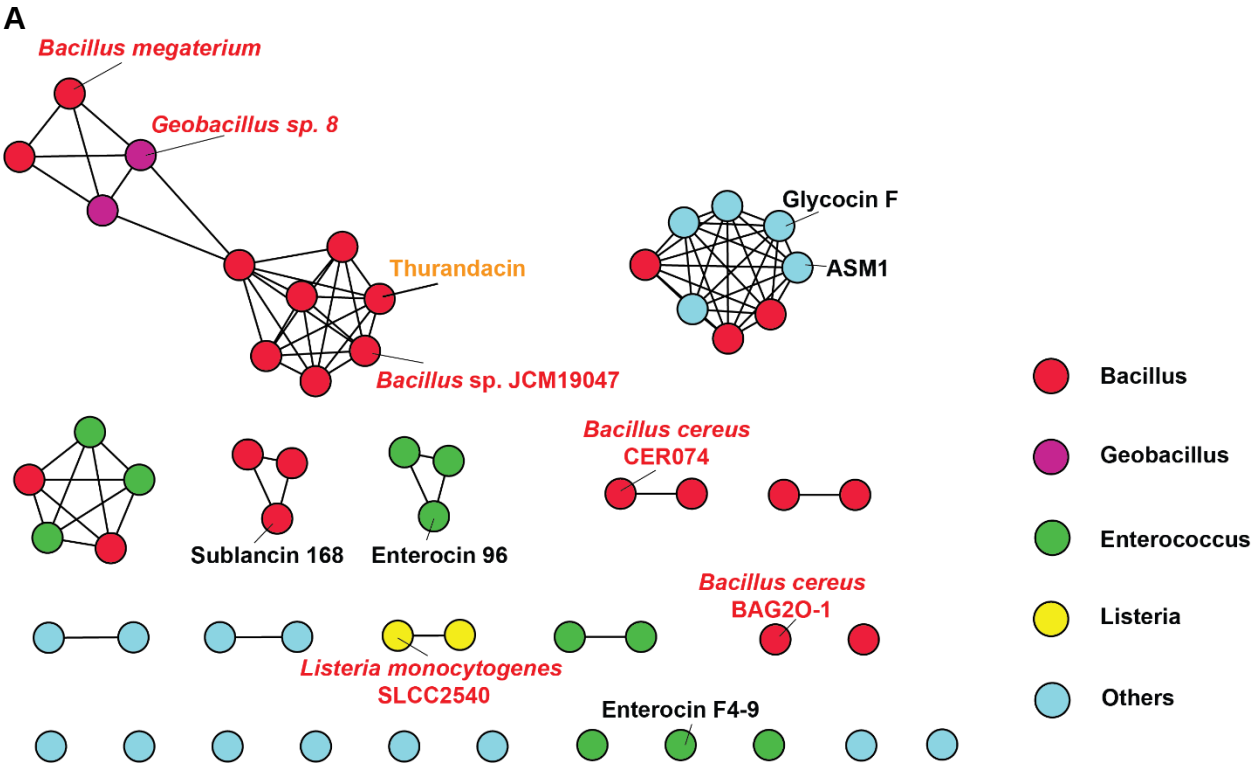
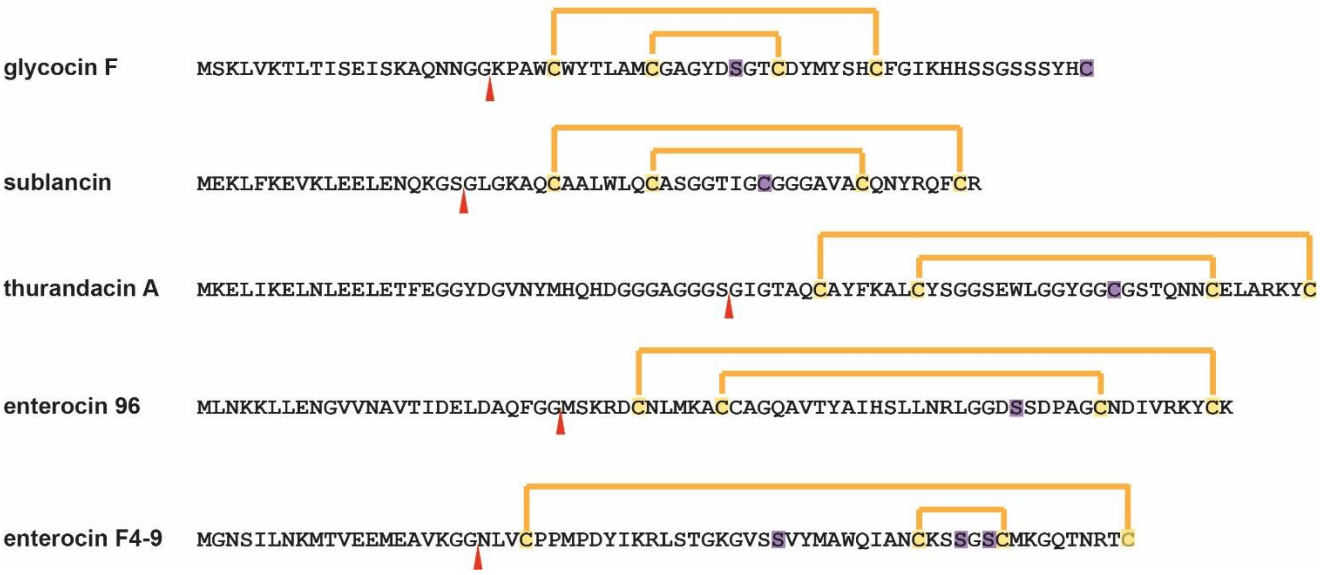


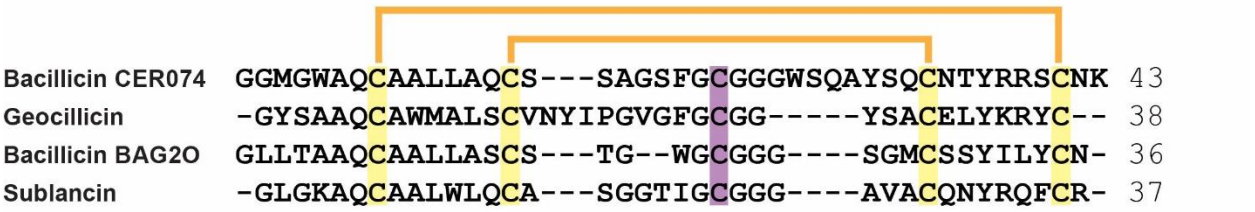
Figure 3.2. Continue

B



C

SunA/ThuA-type



Enterocin 96-type

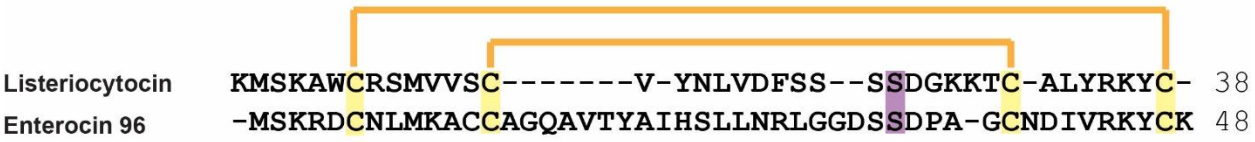


Figure 3.3. A) The selected putative glycocin BGCs. B) Alignment of the putative precursor peptides to the precursor peptide of sublancin, a known glycocin previously discovered from *Bacillus subtilis* 168. The red arrow indicates the sites where leader peptides are predicted to be cleaved. Cysteines that are predicted to form disulfide bonds are highlighted in yellow. Residues that are predicted to be glycosylated are highlighted in purple. The mutated residues of the precursor peptide from *Listeria monocytogenes* SLCC2540 BGC to prevent proteolysis are highlighted in green. Both of the residues were mutated to alanines. This figure was adapted from (27).

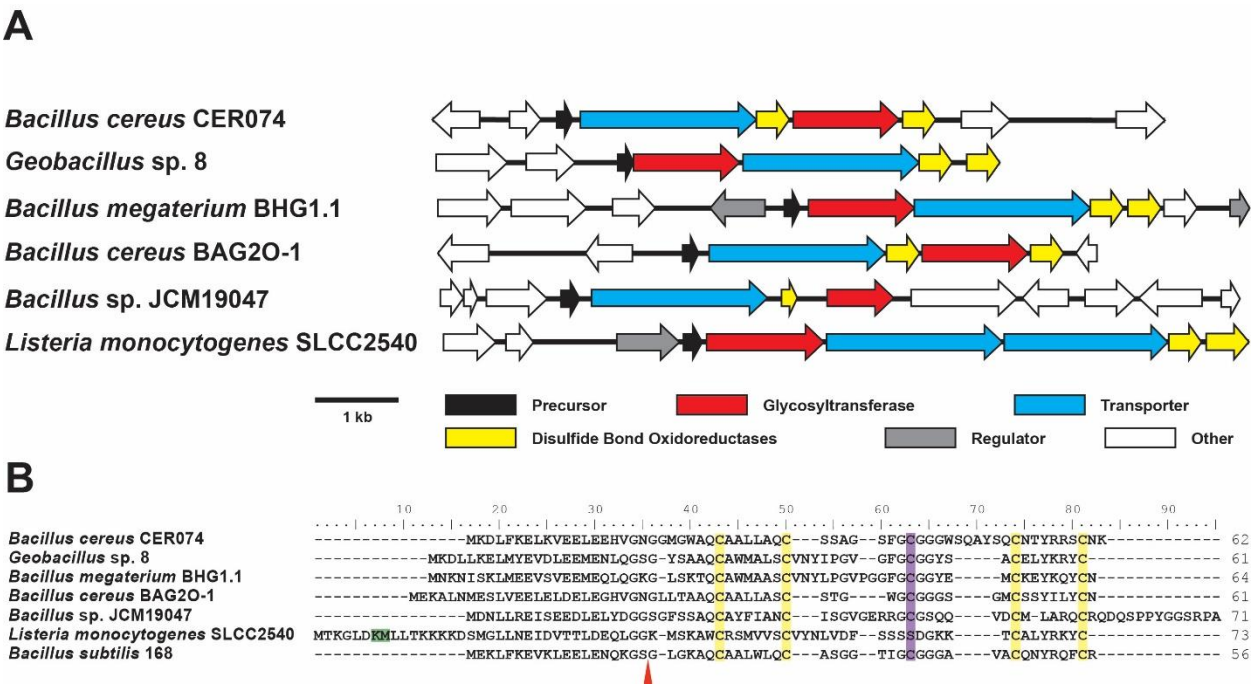


Figure 3.4. The MALDI-MS results of crude extracts from *E. coli* with partially refactored glycocin BGCs. In addition to target MS peaks, other peaks (possibly endogenous peptides from *E. coli*) are also observed. Thus, the result from *E. coli* with empty pET28a plasmid serves as the negative control (MALDI-MS results of all IMAC purified glycosylated precursor peptides are available in Figure 3.5). A) Results from partial BGCs of bacillicin CER074, geocillicin, bacillicin BAG2O and empty pET28a plasmid. Peaks representing the glycosylated precursor peptides are highlighted in yellow. Two new peaks were observed from bacillicin CER074 and the one with higher mass should represent an acetonitrile adduct. B) Results from the partial BGC of listeriocytocin (wild type), listeriocytocin (mutated precursor) and empty pET28a. Peaks representing the N-terminal truncated glycosylated precursor peptide and the glycosylated precursor peptide mutant are highlighted in yellow.

Figure 3.4. Continue

A

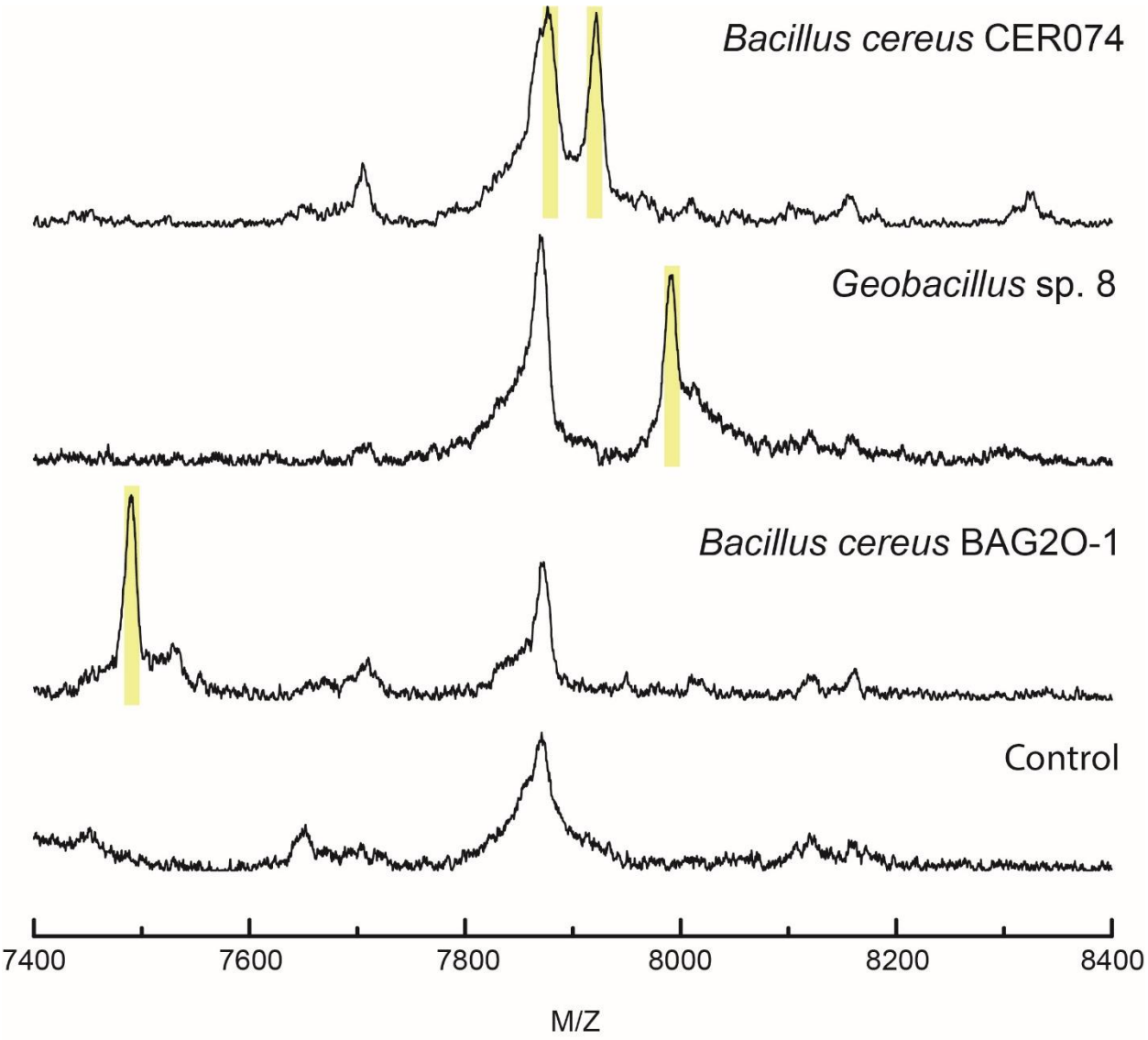


Figure 3.4. Continue

B

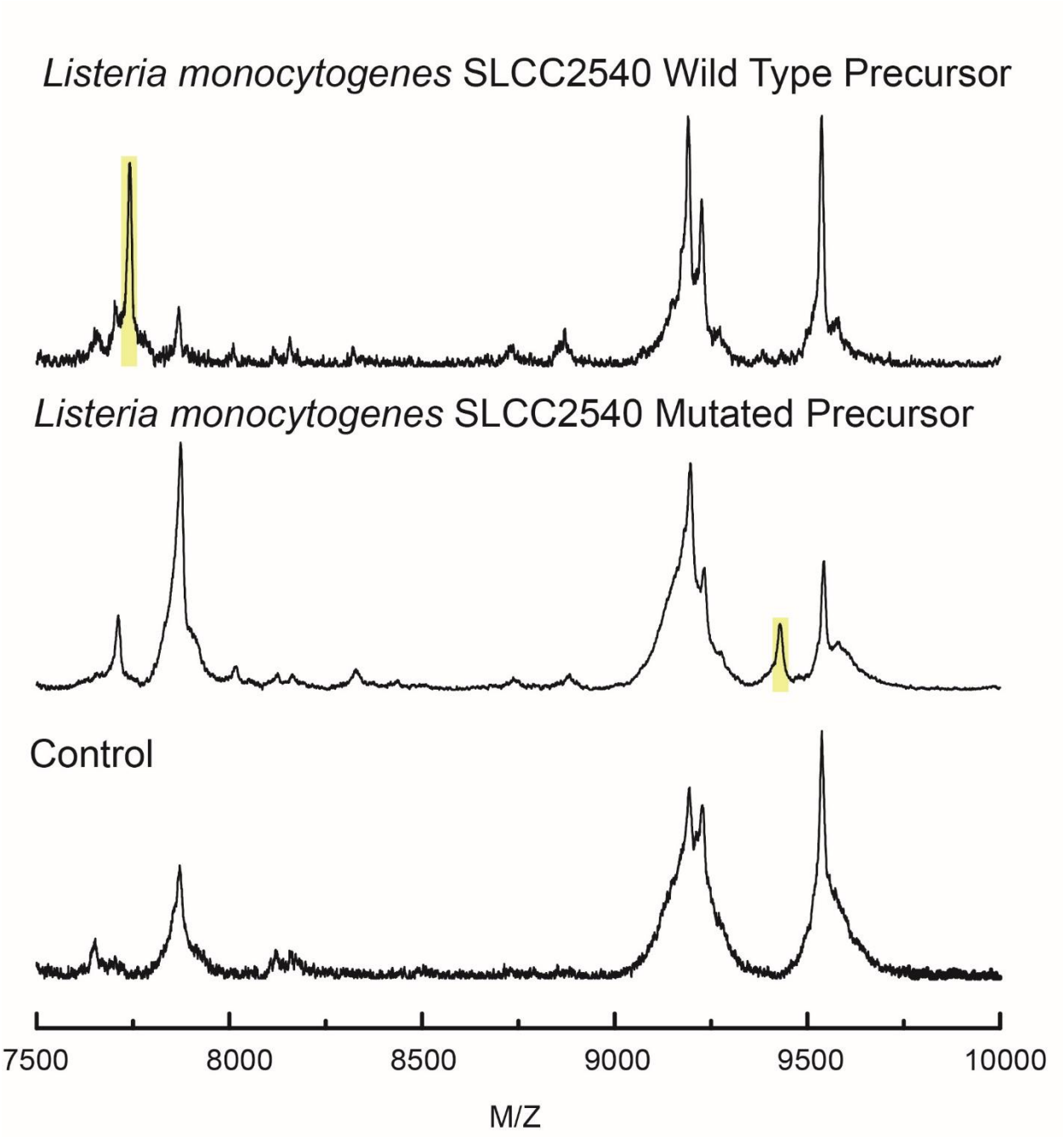


Figure 3.5. MALDI-TOF MS results of the glycosylated precursor peptides purified by IMAC. Bacillicin CER074 intermediate: observed 7882.235 Da, expected 7882.637 Da; geocillicin intermediate: observed 7992.720 Da, expected 7992.098 Da; bacillicin BAG2O intermediate: observed 7490.054 Da, expected 7490.408 Da; listeriocytocin intermediate: observed 9426.555 Da, expected 9426.858 Da. An acetonitrile adduct $[M+H+41]^+$ was also observed in the spectrum of bacillicin CER074 intermediate.

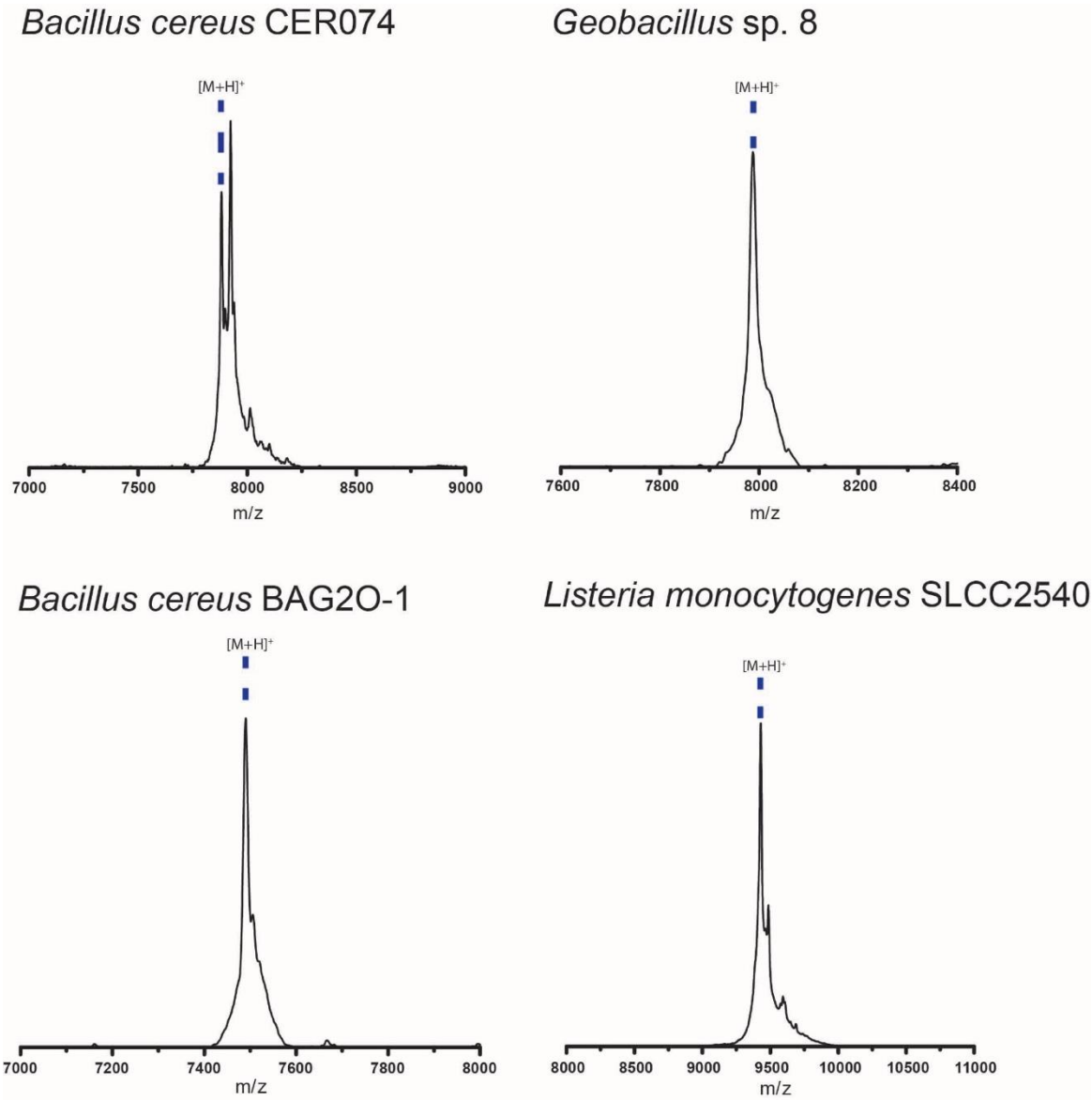


Figure 3.6. MALDI-TOF MS results of M9 liquid culture supernatants of *E. coli* expressing fully refactored glycocin BGCs. The MS peaks of glycocins are highlighted in yellow. The result of *E. coli* with empty pET28a plasmid serves as the negative control.

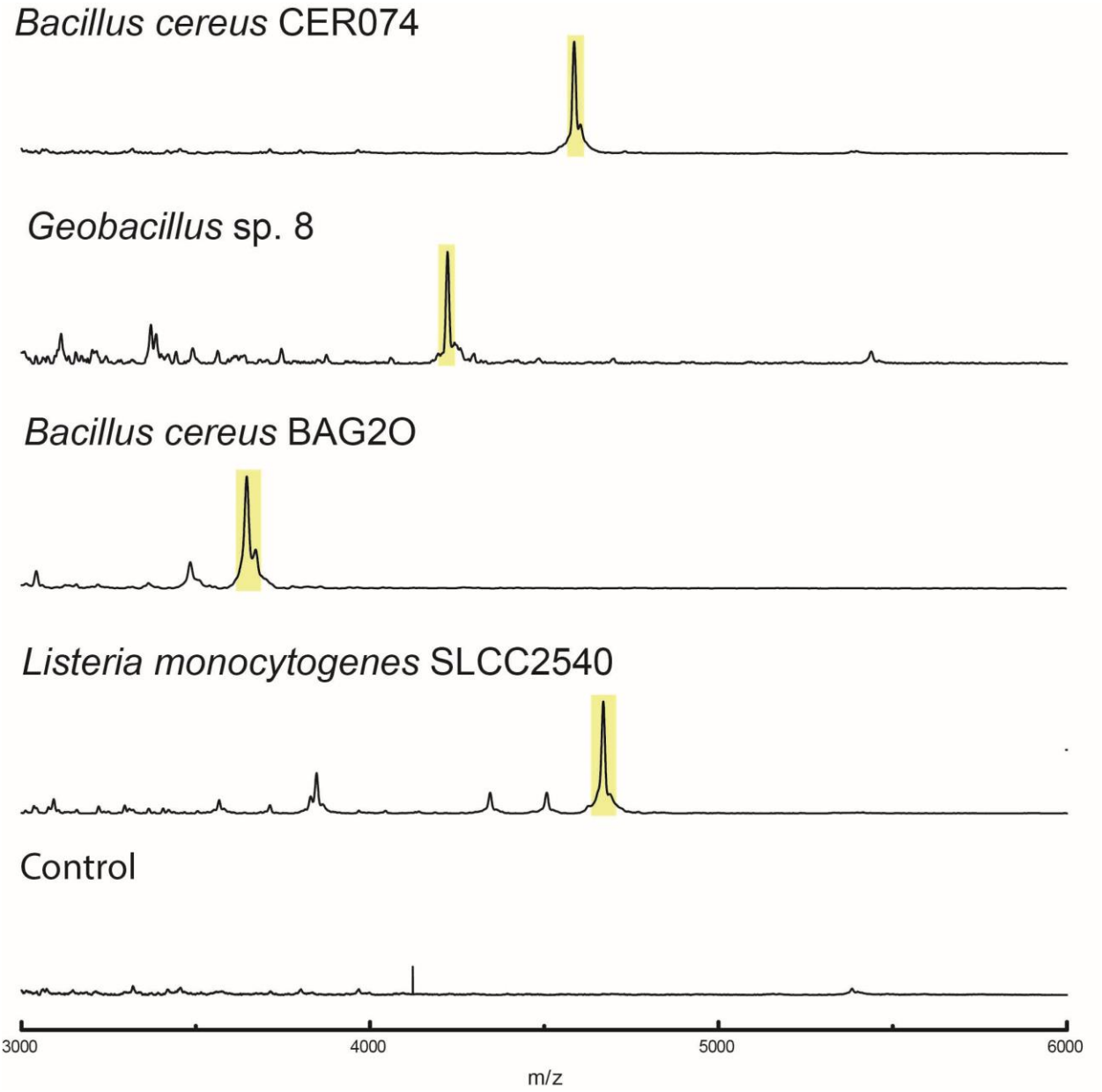


Figure 3.7. Structures of the identified glycocins. A) Structures of the three SunA/ThuA type glycocins discovered in this work. Disulfide bonds are labeled in yellow while the glycosylated cysteines are labeled in purple. B) The 220 nm trace of the analytical HPLC purification. Peaks containing target glycocins are highlighted in yellow. C) MALDI-TOF MS analysis of the purified glycocins (bacillicin CER074: observed 4584.861 Da, expected 4584.057 Da; geocillicin: observed 4223.228 Da, expected 4222.822 Da; bacillicin BAG2O: observed 3650.761 Da, expected 3650.159 Da; listeriocytocin: observed 4667.945 Da, expected 4667.390 Da). D) GC traces of the hydrolyzed and derivatized glycosylated peptides. Traces for multiple hexose standards (galactose, glucose and mannose) were obtained in the same procedure and shown for comparison. This figure was adapted from (27). Figure 3.7D made by Subhanip Biswas.

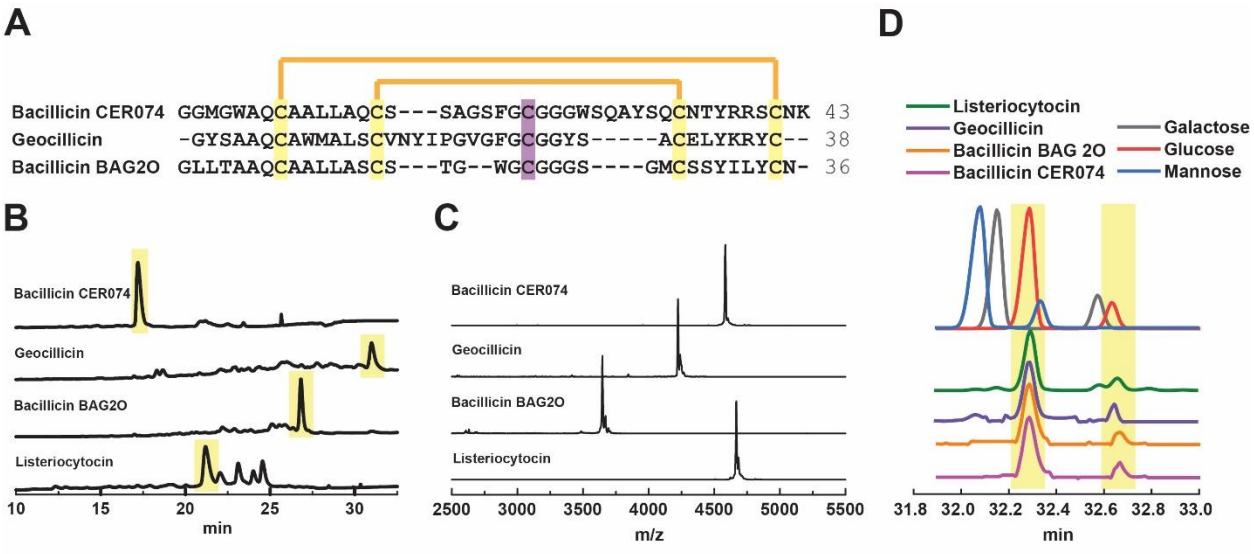


Figure 3.8. Iodoacetamide (IAA) assay of purified glycocins. Results under non-reducing condition are shown in black. Results under reducing condition are shown in red.

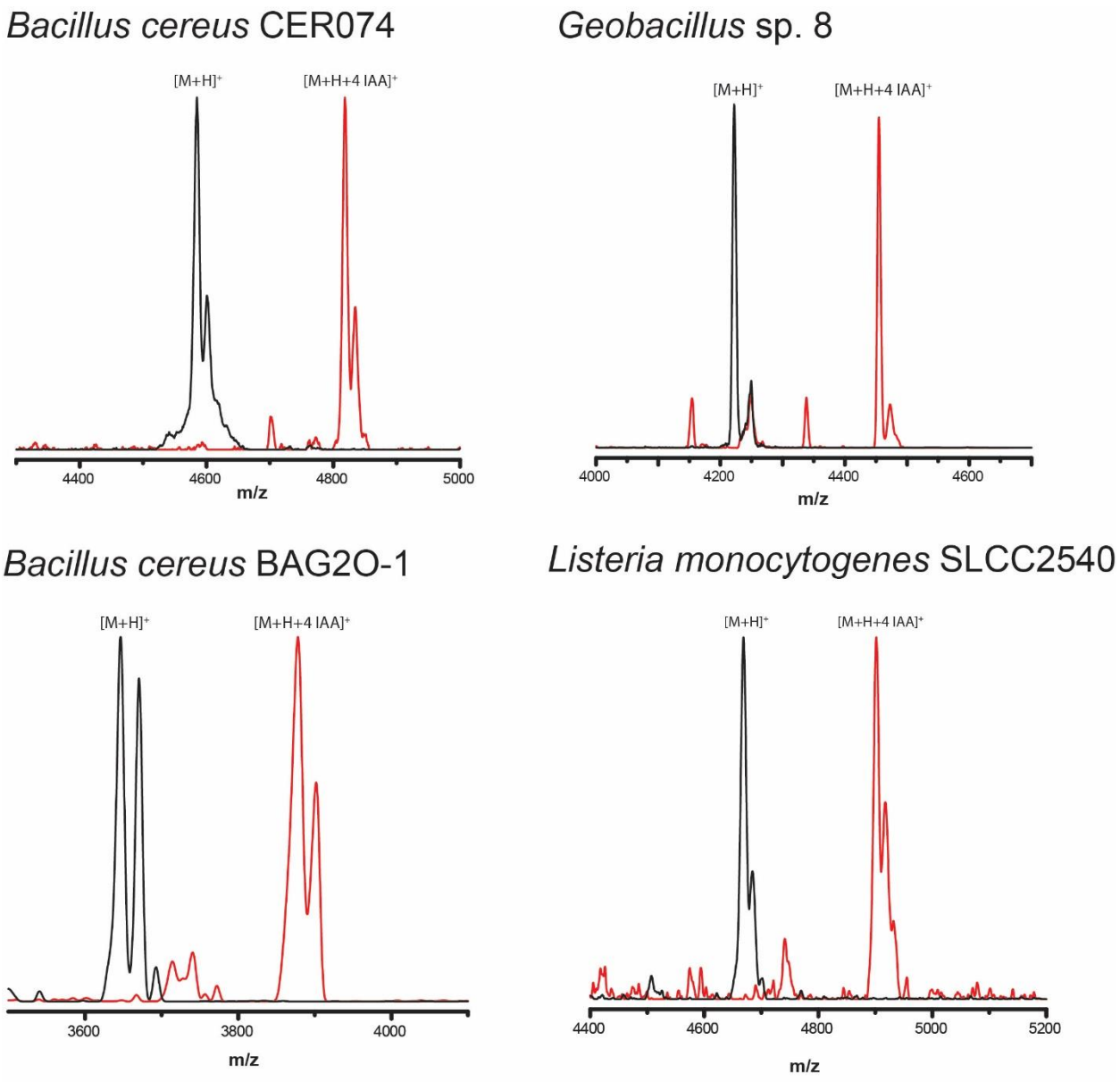


Figure 3.9. Determination of the topology of purified glycocins. All peptides were treated with chymotrypsin under reducing or non-reducing conditions. The anticipated proteolytic fragments under these conditions are color coded in both the structures and mass spectra. Figure made by Subhanip Biswas.

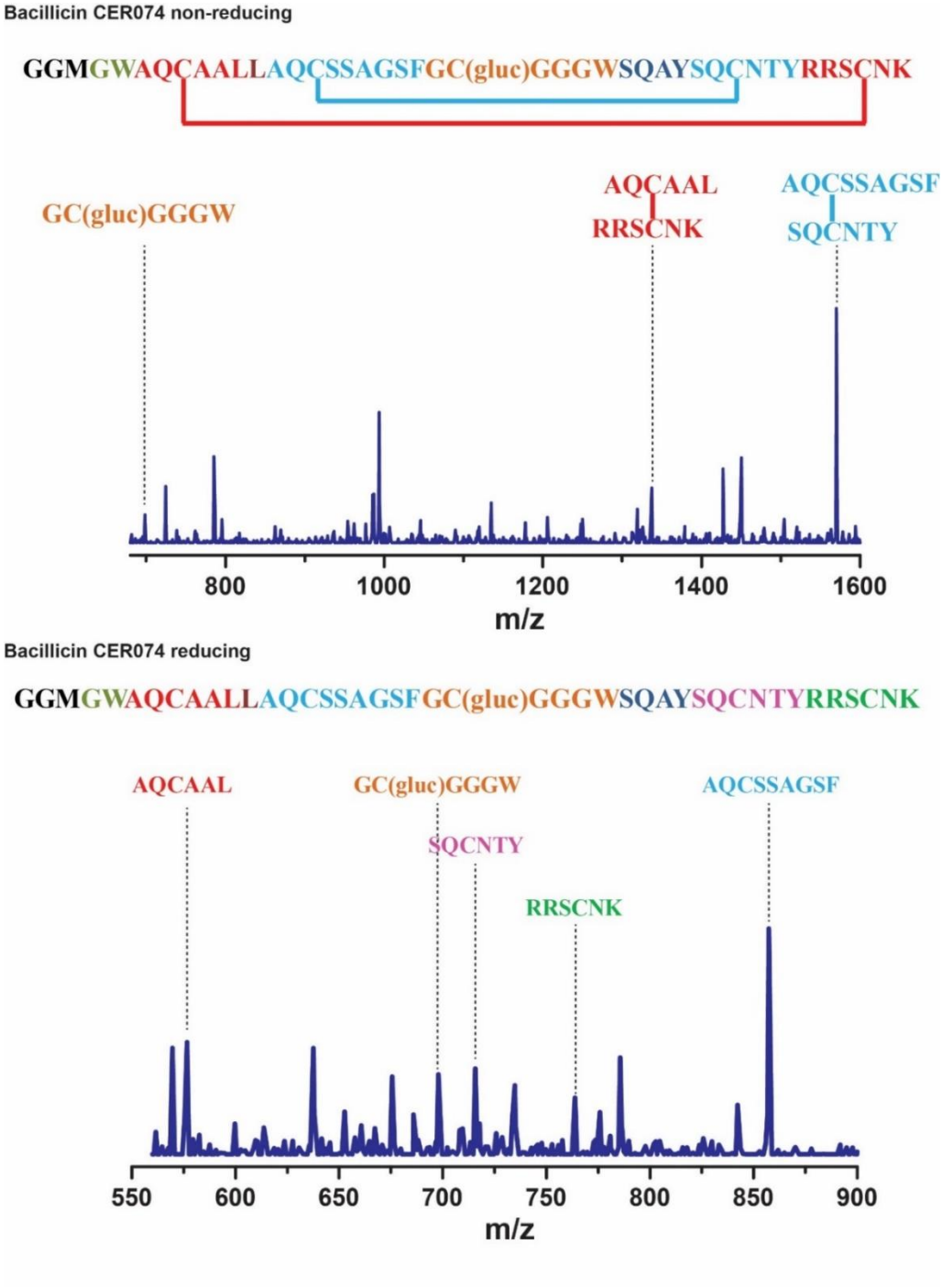


Figure 3.9. Continue

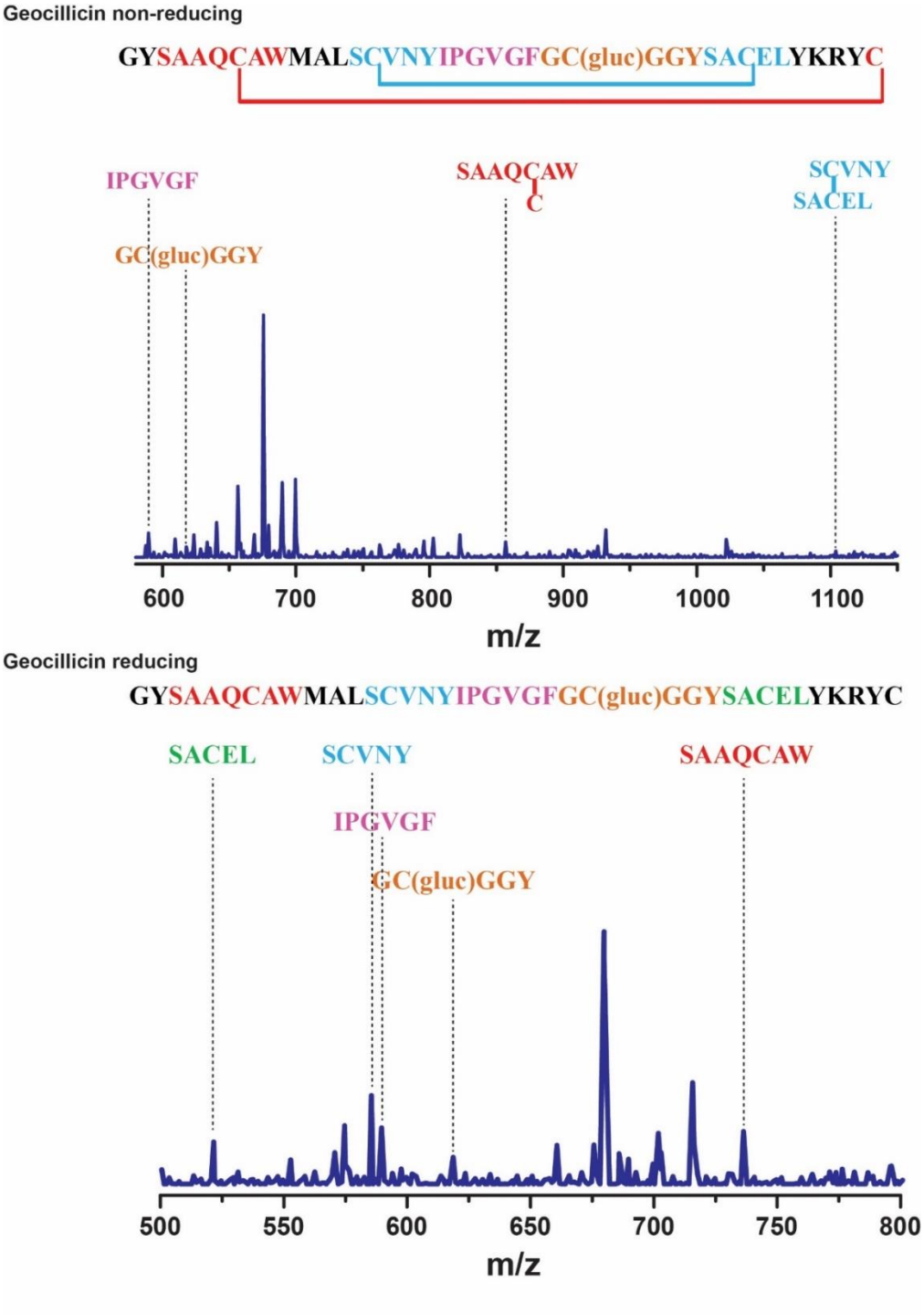


Figure 3.9. Continue

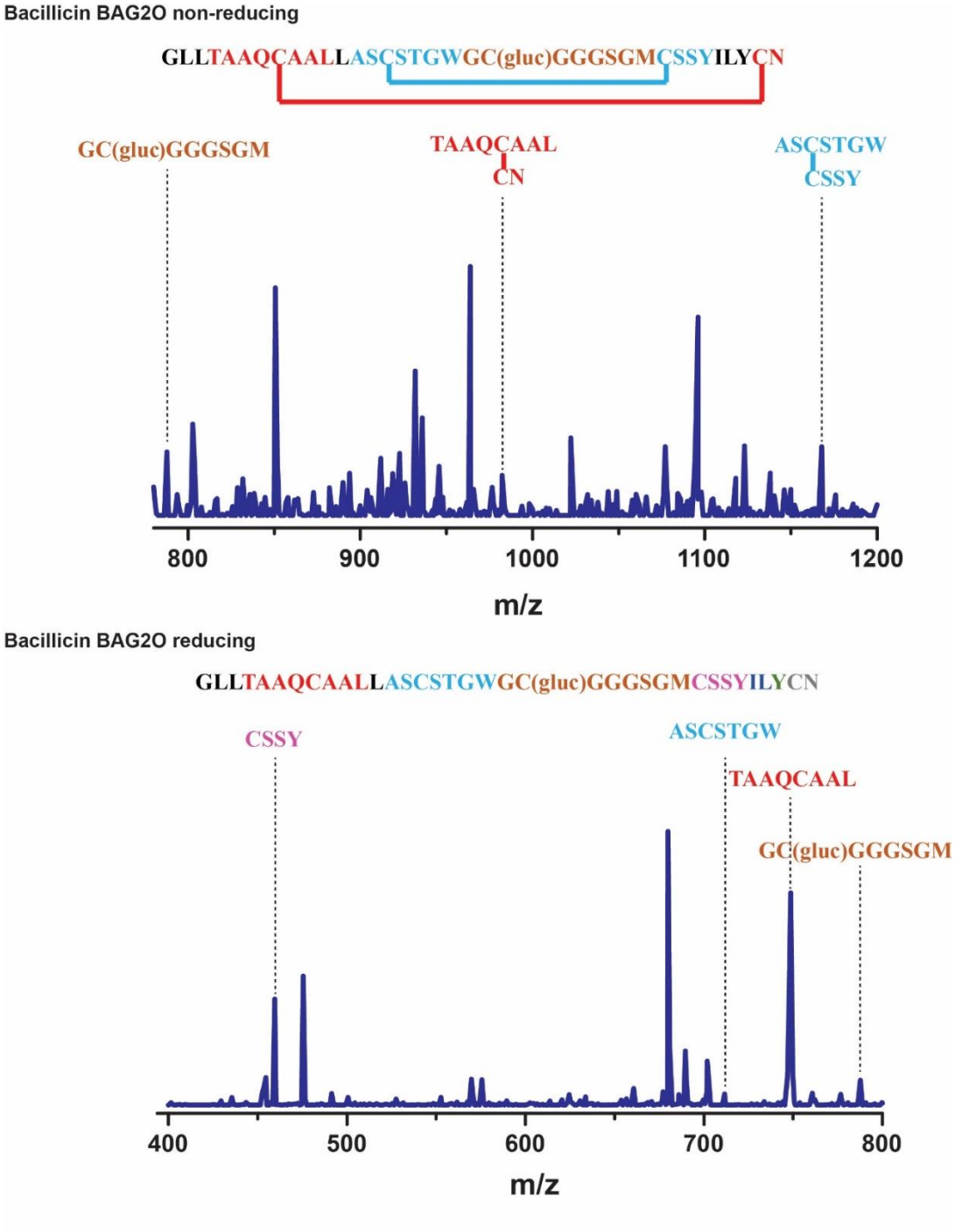


Figure 3.9. Continue

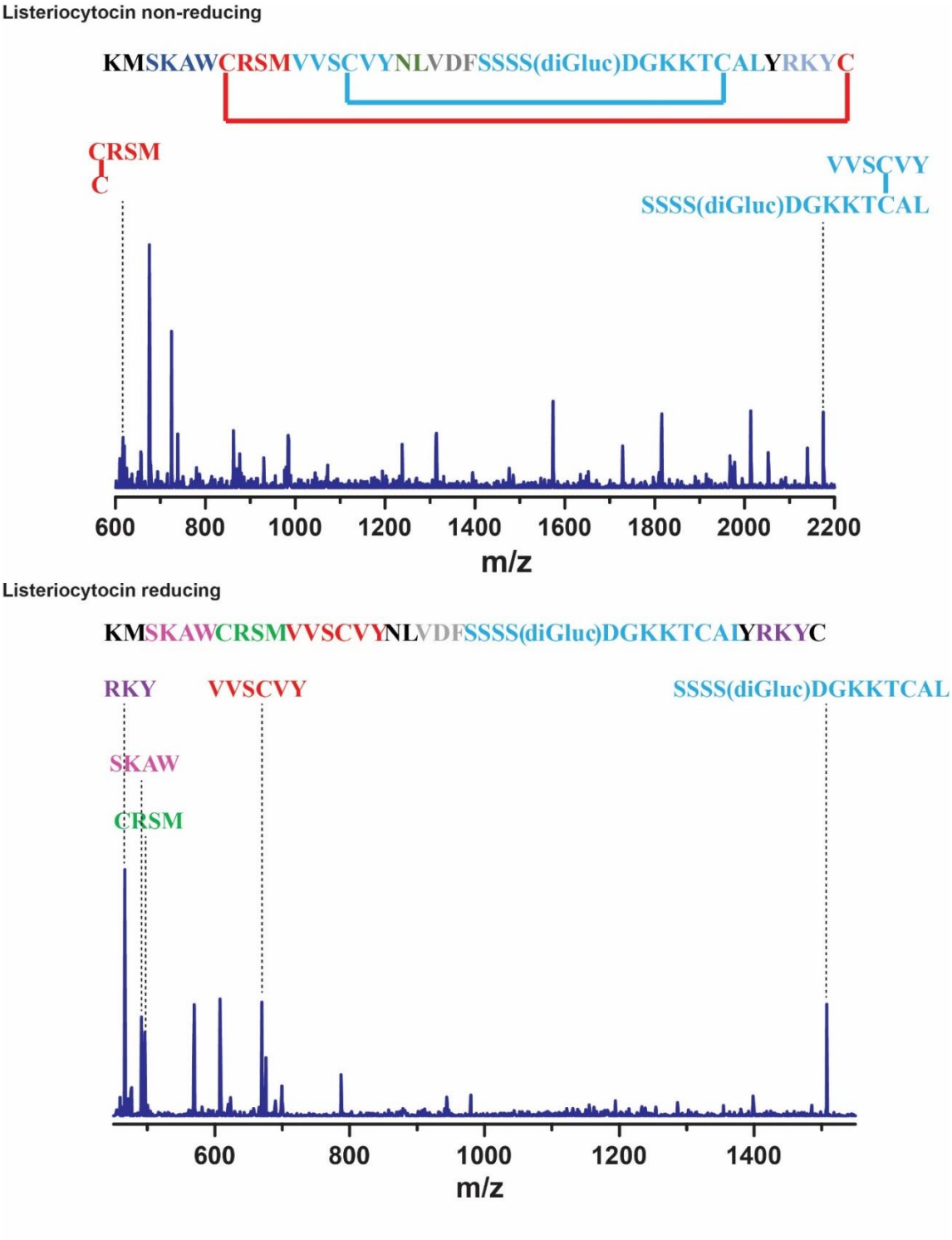


Figure 3.10. Structure of listeriocytocin and MS/MS analysis to determine the site(s) of glycosylation. Cysteine residues that form disulfide bonds are highlighted in yellow and the diglycosylated serine residue is highlighted in purple. The target fragment (underlined with dashed line) for MS/MS analysis was generated by Asp-N endoproteinase under reducing condition. Dehydrated b-ions (-18 Da) are also labeled in the spectrum. This figure was adapted from (27).

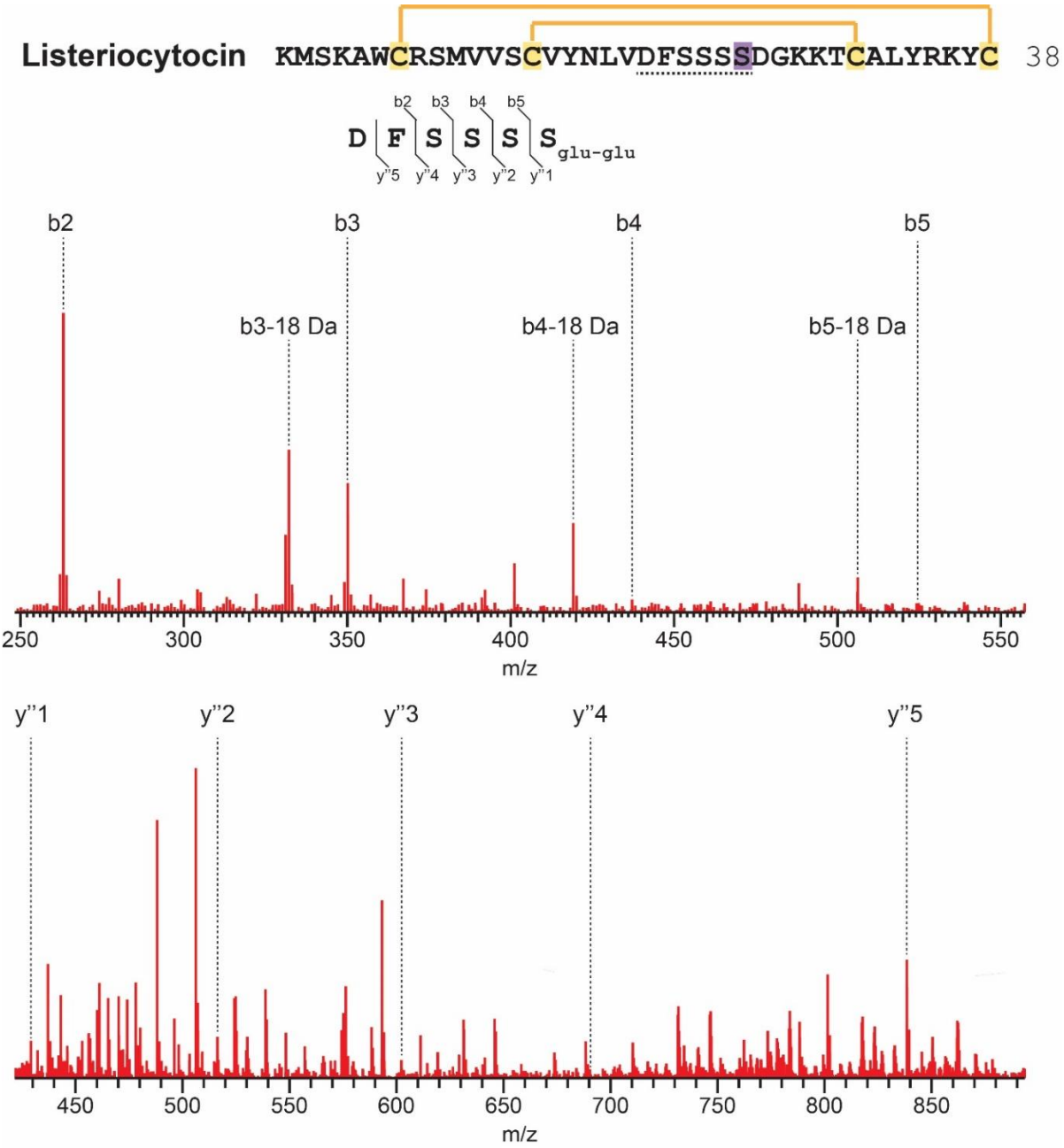


Figure 3.11. MS/MS analysis for an endoproteinase Lys-N fragment of reduced listeriocytocin.

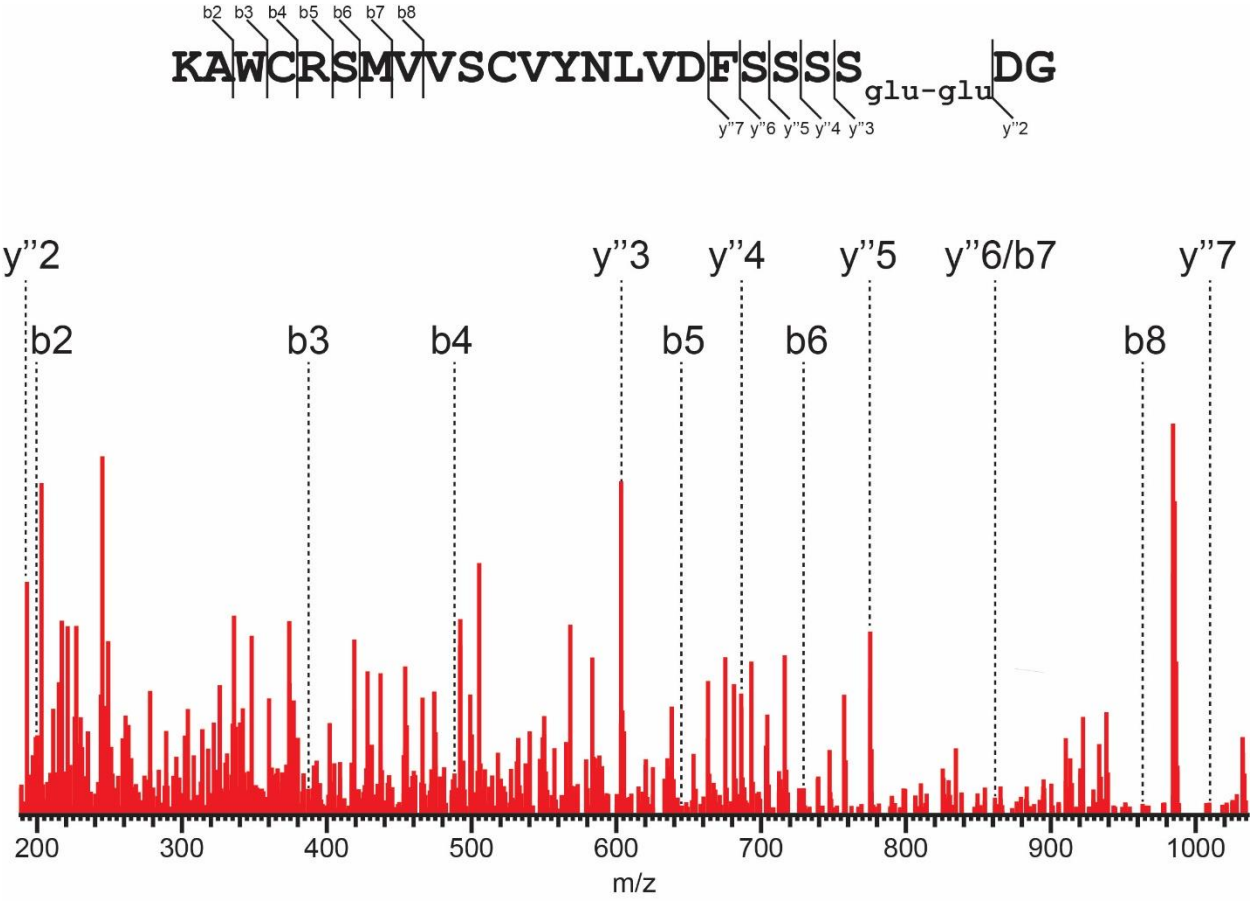


Figure 3.12. MS/MS analysis for the endoproteinase Lys-C fragment of reduced listeriocytocin.

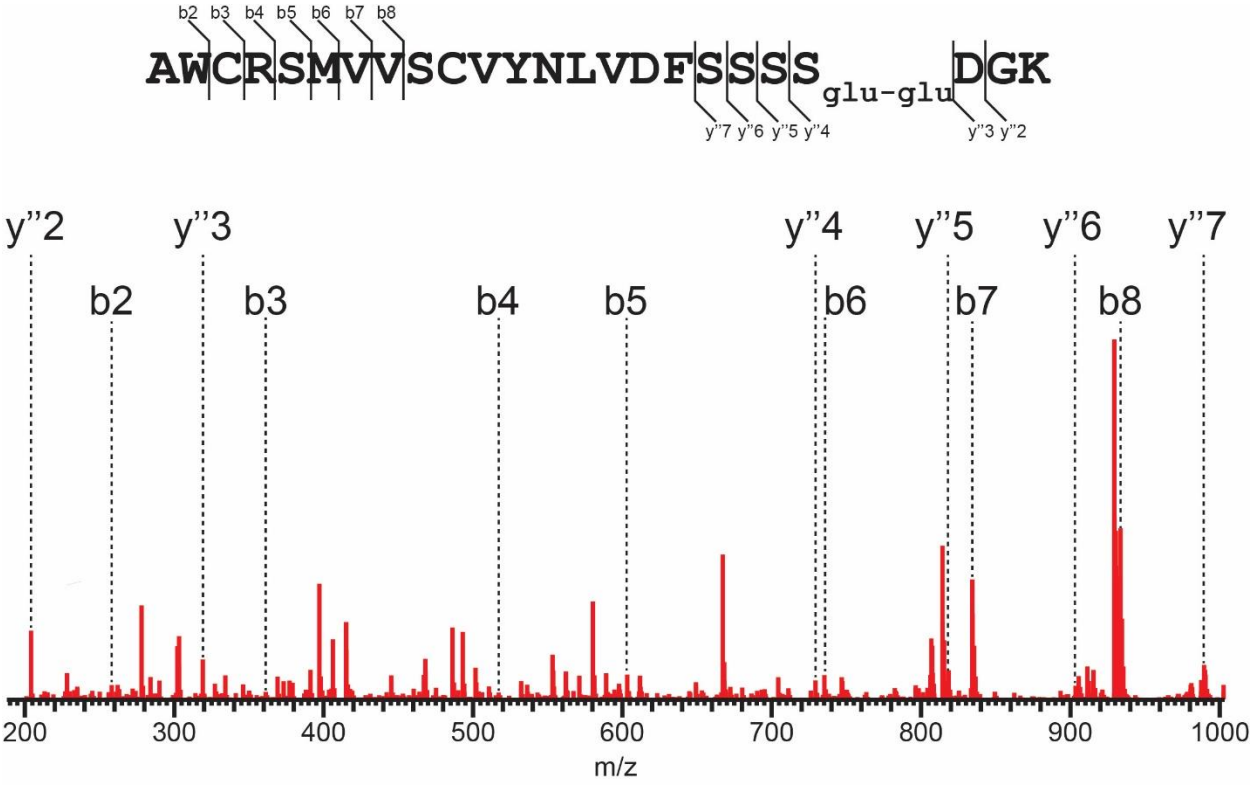


Figure 3.13. Design of the cloning site of T7 His Helper 1.

Figure 3.14. Coexpression of the precursor peptide of enterocin F4-9 with the corresponding glycosyltransferase. A) The amino acid sequence of the precursor peptide of enterocin F4-9 with N-terminal his-tag. B) The MALDI-MS results from the coexpression of enterocin F4-9 precursor peptide and the glycosyltransferase in BL21(DE3) under 37 ºC with 0.5 mM IPTG (observed: 8665.977 Da, expected: 8665.991 Da). BL21(DE3) with empty pET28a plasmid was used as negative control.

A

N-terminal his-tagged Enterocin F4-9 precursor peptide:

MHHHHHHAMGNSILNKMTVEEMEAVKGGNLVCPPMPDYIKRLSTGKGVSSVYMAWQ
IANCKSSGSCMKGQTNRTC

B

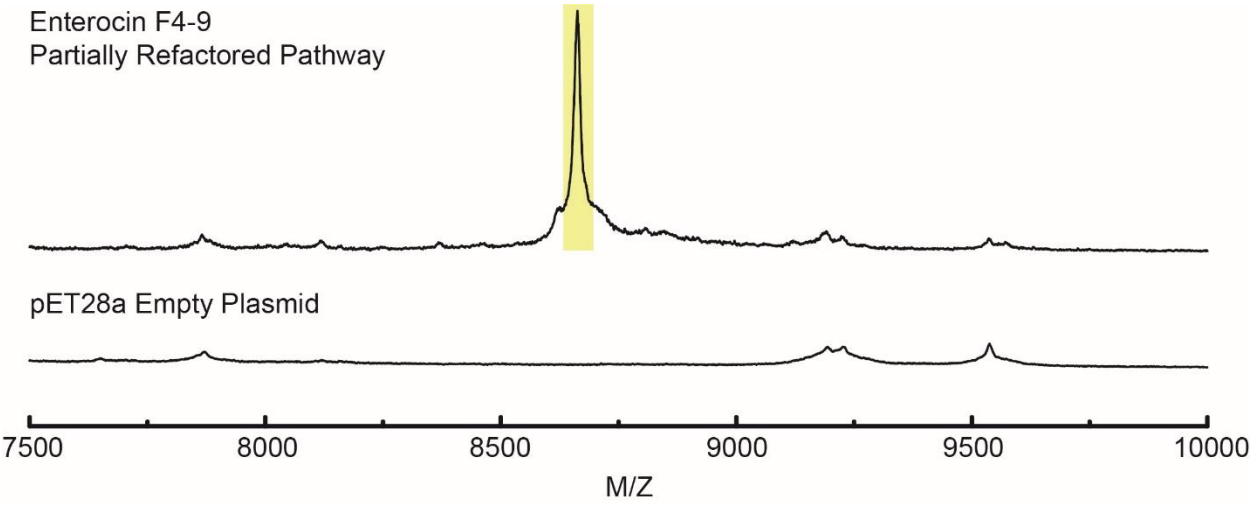


Figure 3.15. Activity of glycocins against *Bacillus cereus* ATCC14579. Samples were spotted on LB agar in a volume of 20 μL. a-e: 20 μM glycocin. f-j: 20 μM glycocin and 500 mM glucose. Glycocins tested: (a,f) sublancin, (b,g) bacillicin CER074, (c,h) bacillicin BAG2O, (d,i) geocillicin, and (e,j) listeriocytocin. k: 500 mM, glucose. l: water. This figure was adapted from (27). Figure made by Subhanip Biswas.

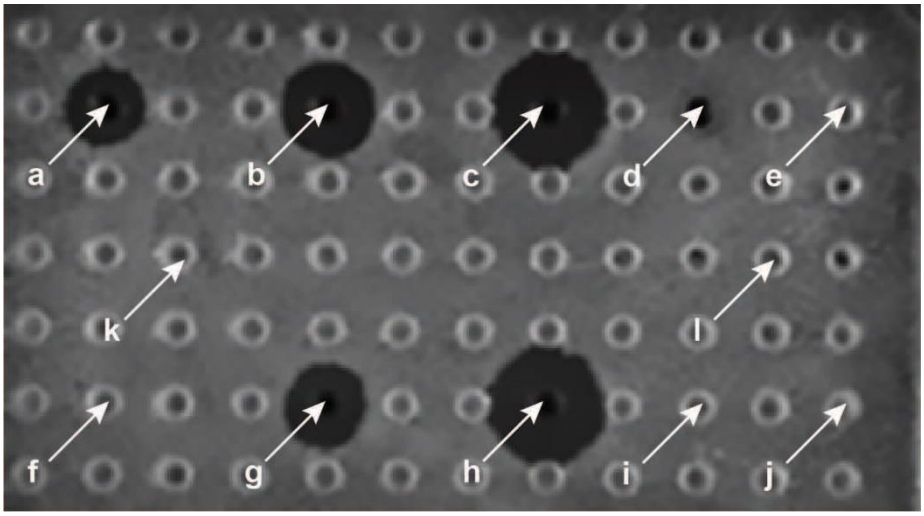


Figure 3.16. Antimicrobial activity assay of glycocins against *Bacillus subtilis* Δspβ. Figure made by Subhanip Biswas.

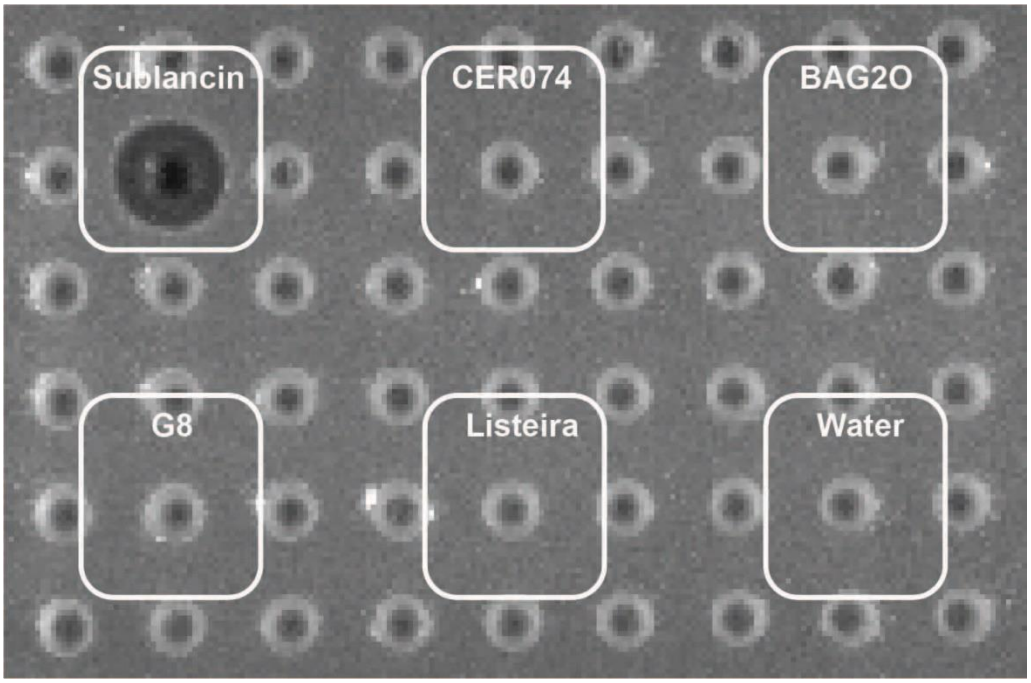


Figure 3.17. Activity assay of glycocins against *Bacillus halodurans* C-125. Figure made by Subhanip Biswas.

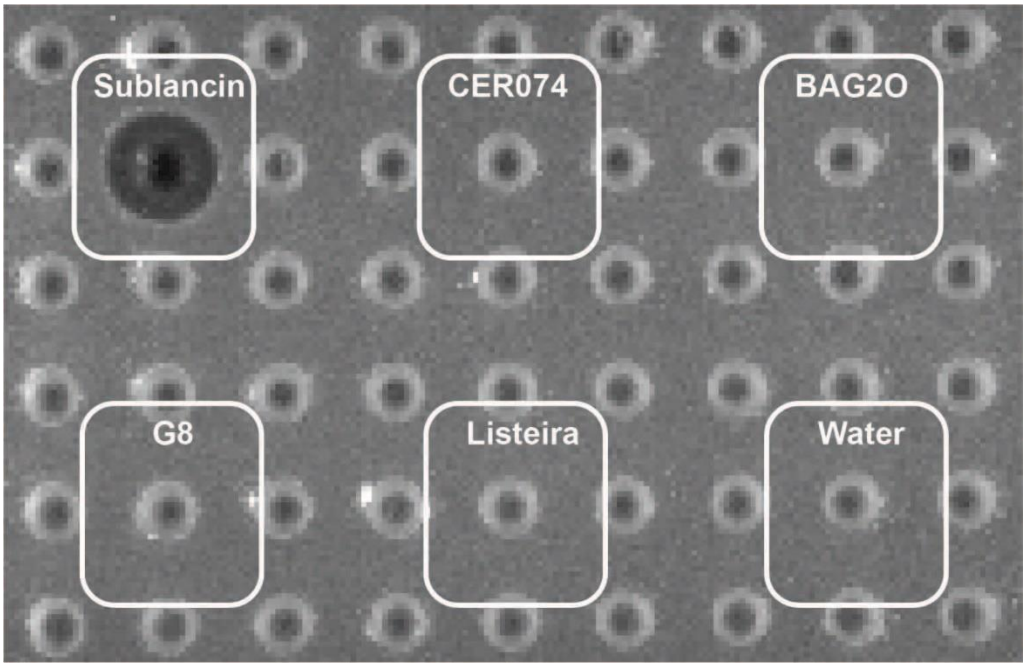


Table 3.1. List of the Tested Indicator Strains

Phylum	Strain
Proteobacteria	*Escherichia coli* JM109
	Pseudomonas aeruginosa PA01
Firmicutes	Methicillin-resistant *Streptococcus aureus* NRS384/USA300
	Vancomycin-resistant *Enterococcus faecium* V583 CB807
	Listeria monocytogenes 4b F2365
	Bacillus subtilis Δspβ
	Bacillus halodurans C-125
	Bacillus cereus ATCC14579

Table 3.2. Accession numbers of the putative precursor peptides and glycosyltransferases for glycocin biosynthesis. A few SunS/SunA like genes have multiple neighboring SunA/SunS like genes.

SunS like	SunA like
WP_083297906.1	WP_070517558.1
WP_095209998.1	WP_048536305.1
WP_095150209.1	WP_038479632.1
WP_065535858.1	WP_010715014.1
WP_065535858.1	WP_002382821.1
WP_065535858.1	WP_010715014.1
WP_065535858.1	WP_002382821.1
WP_094045488.1	WP_094045492.1
WP_090978893.1	WP_090978896.1
WP_090441720.1	WP_090441715.1
WP_074562075.1	WP_074562076.1
WP_070308993.1	WP_070308991.1
WP_038809128.1	WP_033647217.1
WP_016629870.1	WP_002414741.1
WP_013643489.1	WP_013643491.1
WP_010829283.1	WP_002414741.1
WP_010819649.1	WP_002382821.1
WP_010819649.1	WP_010715014.1
WP_010715722.1	WP_002382821.1
WP_010715722.1	WP_010715014.1
WP_010715013.1	WP_002382821.1
WP_010715013.1	WP_010715014.1
WP_002405596.1	WP_010715014.1

Table 3.2. Continue

WP_002405596.1	WP_002405597.1
WP_000083335.1	WP_000183241.1
WP_032089701.1	WP_032089702.1
WP_002382823.1	WP_002382821.1
WP_002382823.1	WP_010715014.1
WP_002368126.1	WP_010706930.1
WP_002368126.1	WP_002368123.1
WP_018375368.1	WP_013643491.1
WP_002400709.1	WP_010715014.1
WP_002400709.1	WP_002382821.1
WP_088266045.1	WP_002130747.1
WP_002130752.1	WP_002130747.1
WP_088040542.1	WP_061530458.1
WP_088013374.1	WP_088013405.1
WP_087305944.1	WP_087305914.1
WP_084167436.1	WP_038479632.1
WP_083637097.1	WP_075601392.1
OQR53364.1	OQR53367.1
WP_081113339.1	WP_061859978.1
WP_055735221.1	WP_038479632.1
WP_078187486.1	WP_078187489.1
WP_076872729.1	WP_048536305.1
OLP15523.1	OLP15521.1
OLP15523.1	OLP15529.1
OKZ96694.1	OKZ96695.1
WP_046648569.1	WP_046648574.1
OFM77885.1	OFM77881.1
AOF43520.1	AOF43525.1
WP_066251544.1	WP_066251537.1
SCH32510.1	SCH32548.1
KZM53254.1	KZM53253.1
WP_061530461.1	WP_061530458.1
KXX95902.1	KXX95899.1
WP_042332806.1	WP_000661240.1
WP_042332806.1	WP_000283312.1
WP_060398790.1	WP_009967544.1
WP_059031426.1	WP_059031428.1
WP_041351120.1	WP_009967544.1
WP_016082809.1	WP_016082806.1
WP_004399050.1	WP_009967544.1

Table 3.2. Continue

WP_060398790.1	WP_009967544.1
GAF22634.1	GAF22631.1
WP_014929842.1	WP_041176876.1
WP_045552786.1	WP_041176876.1
WP_035396167.1	WP_035396166.1
AIC94359.1	AIC94358.1
ETC92743.1	ETC92737.1
WP_006929567.1	WP_006929572.1
WP_016135757.1	WP_016135756.1
WP_015133270.1	WP_015133271.1
WP_002396201.1	WP_002396198.1
EJS79924.1	EJS79923.1
EJS79924.1	EJS79929.1
ADV57361.1	ADV57366.1
EEM68351.1	EEM68348.1
EEM68351.1	EEM68353.1

Table 3.3. The adaptor information for the T7 helper plasmids built in this study.

Position	Adaptor (Left)	Adaptor (Right)
4	CGGT	GTGC
5	GTGC	GACT
2/9	CCCT	AGCG
5/9	GTGC	AGCG
6/9	GACT	AGCG

3.6 References

1. Arnison PG, *et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30(12):1568-1568.
2. Norris GE & Patchett ML (2016) The glycocins: in a class of their own. *Curr Opin Struct Biol* 40:112-119.
3. Stein T (2005) Bacillus subtilis antibiotics: structures, syntheses and specific functions. *Mol. Microbiol.* 56(4):845-857.
4. Wang S, *et al.* (2017) Use of the antimicrobial peptide sublancin with combined antibacterial and immunomodulatory activities to protect against methicillin-resistant Staphylococcus aureus infection in mice. *J. Agric. Food. Chem.* 65(39):8595-8605.
5. Wang S, *et al.* (2015) The antimicrobial peptide sublancin ameliorates necrotic enteritis induced by *Clostridium perfringens* in broilers. *J Anim Sci* 93(10):4750-4760.
6. Paik SH, Chakicherla A, & Hansen JN (1998) Identification and characterization of the structural and transporter genes for, and the chemical and biological properties of, sublancin 168, a novel lantibiotic produced by *Bacillus subtilis* 168. *J. Biol. Chem.* 273(36):23134-23142.
7. Ren H, Hu P, & Zhao H (2017) A plug-and-play pathway refactoring workflow for natural product research in *Escherichia coli* and *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 114(8):1847-1854.
8. Tietz JI, *et al.* (2017) A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat Chem Biol* 13(5):470-478.
9. Biswas S, De Gonzalo CVG, Repka LM, & van der Donk WA (2017) Structure-activity relationships of the S-Linked glycocin sublancin. *ACS Chem Biol* 12(12):2965-2969.
10. Kaunietis A, de Jong A, Pranckute R, Buivydas A, & Kuipers OP (2016) Draft genome sequences of two geobacillus species strains, isolated from oil wells and surface soil above oil pools. *Genome Announc* 4(5):01129-01116.
11. Kudo T, *et al.* (2014) Draft genome sequences of cyclodextrin-producing alkaliphilic bacillus strains JCM 19045, JCM 19046, and JCM 19047. *Genome Announc* 2(2):e00211-00214.
12. Kuenne C, *et al.* (2013) Reassessment of the *Listeria monocytogenes* pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. *BMC Genomics* 14:47.
13. Van der Auwera GA, Feldgarden M, Kolter R, & Mahillon J (2013) Whole-genome sequences of 94 environmental isolates of *Bacillus cereus Sensu Lato*. *Genome Announc* 1(5):e00380-00313.
14. Wang W, *et al.* (2016) Draft genome sequence of *Bacillus megaterium* BHG1.1, a strain isolated from bar-headed goose (*Anser indicus*) feces on the Qinghai-Tibet plateau. *Genome Announc* 4(3):e00317-00316.
15. Oman TJ, Boettcher JM, Wang HA, Okalibe XN, & van der Donk WA (2011) Sublancin is not a lantibiotic but an S-linked glycopeptide. *Nat Chem Biol* 7(2):78-80.
16. Nagar R & Rao A (2017) An iterative glycosyltransferase EntS catalyzes transfer and extension of O- and S-linked monosaccharide in enterocin 96. *Glycobiology* 27(8):766-776.

17. Guo AC, *et al.* (2013) ECMDB: the *E. coli* metabolome database. *Nucleic Acids Res.* 41(D1):625-630.
18. Sajed T, *et al.* (2016) ECMDB 2.0: a richer resource for understanding the biochemistry of E. coli. *Nucleic Acids Res.* 44(D1):495-501.
19. Izquierdo E, Wagner C, Marchioni E, Aoude-Werner D, & Ennahar S (2009) Enterocin 96, a novel class II bacteriocin produced by *Enterococcus faecalis* WHE 96, isolated from munster cheese. *Appl Environ Microbiol* 75(13):4273-4276.
20. Maky MA, *et al.* (2015) Enterocin F4-9, a novel O-linked glycosylated bacteriocin. *Appl Environ Microbiol* 81(14):4819-4826.
21. Stepper J, *et al.* (2011) Cysteine S-glycosylation, a new post-translational modification found in glycopeptide bacteriocins. *FEBS Lett.* 585(4):645-650.
22. Bisset SW, *et al.* (2018) Using chemical synthesis to probe structure-activity relationships of the glycoactive bacteriocin glycocin F. *ACS Chem Biol* 13(5):1270-1278.
23. De Gonzalo CVG, *et al.* (2015) The phosphoenolpyruvate:sugar phosphotransferase system Is involved in sensitivity to the glucosylated bacteriocin sublancin. *Antimicrob Agents Chemother.* 59(11):6844-6854.
24. Kouwen TRHM, *et al.* (2009) The large mechanosensitive channel MscL determines bacterial susceptibility to the bacteriocin sublancin 168. *Antimicrob Agents Chemother.* 53(11):4702-4711.
25. Neidhardt FC, Bloch PL, & Smith DF (1974) Culture medium for enterobacteria. *J Bacteriol* 119(3):736-747.
26. Li B, Cooper LE, & van der Donk WA (2009) *In vitro* studies of lantibiotic biosynthesis. *Complex Enzymes in Microbial Natural Product Biosynthesis, Part A: Overview Articles and Peptides* 458:533-558.
27. Ren H, Biswas S, Ho S, van der Donk WA, & Zhao HM (2018) Rapid discovery of glycocins through pathway refactoring in *Escherichia coli*. *ACS Chem Biol* 13(10):2966-2972.

Chapter 4: Discovery and Characterization of a Class IV Lanthipeptide with Novel Ring Topology

4.1 Introduction

Lanthipeptides, which is short for lanthionine-containing peptides, constitute a major family of ribosomally synthesized and post-translationally modified peptides (RiPPs) (1). Historically, lanthipeptides were named as lantibiotics due to their antimicrobial activity. Nisin, the founding member of the lanthipeptide class of natural products, was discovered from *Lactococcus lactis* in 1928 as a substance that inhibits the growth of *Lactobacillus bulgaricus* (2). The subsequent demonstration of its broad spectrum antimicrobial activity resulted in its use in food preservation till today. However, since the beginning of the 21st century, the discovery rate of lanthipeptides is significantly increased due to technical advances in DNA sequencing and development of bioinformatics tools for biosynthetic gene cluster prediction and researchers began to realize that the activity of lanthipeptides is not restricted to antimicrobial activity (3, 4). More than a hundred lanthipeptides are known to date which exhibit very diverse activities including antifungal (5), morphogenetic (6, 7), antiviral (8), antinociceptive (9), and antiallodynic functions (10). More interestingly, lanthipeptide derivatives are undergoing therapeutic evaluation and have been used for imaging applications (11).

The biosynthesis of lanthipeptides resemble most of RiPP natural products (11). The precursor peptide is first translated from a small open reading frame (ORF) with the leader peptide region and core peptide region at its N-terminal half and C-terminal half respectively. The leader peptide region is recognized by the corresponding lanthipeptide synthase while modifications were taken place at the core peptide region. The mature lanthipeptide was released by removing the leader

peptide by peptidase(s) encoded either in the pathway or elsewhere in the genome of the producer. The defining structural feature of lanthipeptides is the presence of thioether-bridged macrocycles, which are composed of (methyl)-lanthionine bis-amino acids. Such a ring is formed in a two-step process. First, either serine or threonine undergoes dehydration through beta-elimination on their side chains, forming dehydroalanine (Dha) or dehydrobutyrine (Dhb) respectively. Then, the nucleophilic attack from the thiol group from the cysteine side chain to the double bond on Dha or Dhb yields the thioether bond through 1,4-conjugate addition. Depending on the direction in which the nucleophilic attack happens, two diastereomers, DL-(methyl)lanthionine and LL-(methyl)lanthionine can be formed (12).

Based on the characteristics of lanthipeptide synthases, lanthipepides are classified into four subfamilies (11, 13). Class I lanthipeptides are formed by two separate synthases. LanB catalyze the dehydration reaction while LanC catalyzes the cyclization reaction. While Class II lanthipeptides are synthesized by a single LanM enzyme which harbors both a dehydratase domain and a cyclase domain. Both Class III and Class IV lanthipeptides are synthesized by enzymes that have three domains: an N-terminal lyase domain, a central kinase domain and a C-terminal cyclase domain (14). The main difference between Class III and Class IV lanthipeptide synthases lies in their cyclase domains. While the cyclase domain from Class IV synthases resembles the LanC and the Class II cyclase domain, the cyclase domain from Class III is more different and lacks the canonical coordination sites for Zn^{2+} ions. In Classes II – IV, the dehydration is realized through ATP-dependent phosphorylation of the serine/threonine hydroxyl group. In Class I, the hydroxyl groups are glutamylated prior to elimination which is accomplished by using glutamylated tRNA as substrate (15, 16).

While more than a hundred lanthipeptides have been discovered to date, only very few of them belong to Class IV lanthipeptide (17). The first known Class IV lanthipeptide, venezuelin, was discovered from *Streptomyces venezuelae* (14). Although a few more Class IV lanthipeptides were discovered later on, all of them are clearly homologs of venezuelin (18, 19). In addition, the biosynthetic mechanism was much less studied for both Class III and Class IV lanthipeptides compared with Class I and Class II lanthipeptides. The in trans activity of the lyase and kinase domains from the venezuelin synthase were reconstituted *in vitro* and residues crucial for the activity of the lyase domain were identified (14, 20). The structure of venezuelin was also characterized by *in vitro* maturation of the precursor peptide combined with the tandem mass spectrometry (MS/MS) method (14). In the follow up studies, the structure of venezuelin was confirmed by producing the mature peptide from Streptomyces (18, 19). More recently, the leader peptide binding mechanism was elucidated *in vitro*. It was shown that the α-helical stretch of residues at the center to N-terminal region of the leader peptide was crucial for the recognition by the LanL, while the kinase domain was responsible for the recognition of the leader peptide (21). The dehydration order was also determined as N-C. Besides the biosynthetic mechanism, the bioactivity of Class IV lanthipeptides is also elusive. Only a moderate inhibitory activity against protein tyrosine phosphate 1B (PTP1B) was ever reported for streptocollin, a Class IV lanthipeptide discovered from *Streptomyces collinus* Tu 365 (18).

In this work, a new group of Class IV lanthipeptide biosynthetic pathways was identified through bioinformatics analysis. One of the representatives from *Streptomyces sp.* NRRL S-1022 was refactored and heterologously expressed in *Escherichia coli*, which generates a lanthipeptide with

two non-overlapped rings, a new topology that distinguishs this group of Class IV lanthipeptide from venezuelin and its homologs. In order to further study the mechanism of ring formation, the LanL was expressed and purified for *in vitro* assays. Through analysis of the intermediates by MS/MS, it was found that there are multiple routes for the ring formation while the cyclase domain exhibits some preference on the C to N direction for cyclization. Overall, we have successfully identified a new Class IV lanthipeptide and studied its biosynthesis *in vitro*, which has expanded our knowledge for the Class IV lanthipeptide structures and biosynthesis.

4.2 Results and Discussion

4.2.1 Pathway Prediction and Analysis

In order to pick a BGC with the potential to produce a Class IV lanthipeptide with unknown structural features, we analyzed the data in the Gene Cluster Family database, where potential natural product biosynthetic gene clusters (NPGCs) predicted from 830 actinobacteria genomes were grouped into gene cluster families (GCFs) through similarities (22). GCF128 which contains 13 gene clusters and was classified into lanthipeptides was chosen for further analysis. Among these 13 gene clusters, only two genes were shown to be highly conserved: one was shown as a Class IV lanthipeptide synthase (LanL) based on conserved domain search in the National Center for Biotechnology Information (NCBI) database while the other one was supposed to be the precursor peptide (LanA). Alignment of the LanAs in GCF128 has shown that two cysteine residues, three threonine residues, and one serine residue in the C-terminal half are conserved (Figure 4.1). Compare with the precursor peptide sequence of venezuelin, which has four cysteine residues, GCF128 clearly represents a new group of Class IV lanthipeptide BGCs which can produce Class IV lanthiepeptides with completely new ring topology.

4.2.2 Pathway Refactoring and Heterologous Expression in *E. coli*

To further characterize the product from GCF128, one of the 13 BGCs which was predicted from the genome of *Streptomyces sp*. NRRL S-1022 was chosen for heterologous expression in *Escherichia coli*. Both the LanA and LanL were refactored by T7 promoter and terminator through the plug-and-play workflow we discussed in Chapter 2 (23). After expression in *E. coli* BL21(DE3) and product purification through immobilized metal affinity chromatography (IMAC), the eluate was analyzed on matrix-assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF-MS) and the m/z corresponding to the precursor peptide with 2-fold dehydration was clearly observed (Figure 4.2).

4.2.3 Product Purification and Structural Characterization

To get more structural information, the eluate from IMAC was further purified by semiprep HPLC. After digested by trypsin and analyzed on MALDI-TOF-MS, most of the peaks correlate to the expected fragments from the linear peptide while one peak correlates to the sum of two C – terminal fragments with 2-fold dehydration, which serves as an evidence that the peptide is cyclized (Figure 4.3). To further confirm that the peptide is cyclized, an N–ethylmaleimide (NEM) based derivatization assay was performed. While the unmodified precursor peptide clearly exhibited a mass shift, there was no mass shift for the modified precursor peptide after NEM treatment, which clearly showed that both of the cysteine residues were involved in the cyclization (Figure 4.2).

In order to elucidate the ring topology (e.g. if the two rings are overlapped or not), the trypsin digested fragments of the modified peptide was analyzed by MS/MS, and the data supports a non-overlapped ring topology (Figure 4.4). To further confirm which residue (three threonine and one serine residue), four LanA mutants (S44A, T49A, T51A and T52A) with each of these four residues mutated to alanine was coexpressed with LanL in *E. coli* respectively. While two fold dehydration was observed for both S44A and T51A, only unmodified peptides were observed for T49A and T51A, which supports that the modified peptide has two non-overlapped rings in which T49 and T51 are involved in the ring formation (Figure 4.5).

Of note, the productivity of the modified precursor peptide is less than 1 ug/L, which largely limits further characterization of the peptide, such as the stereochemistry of its thioether crosslinks and bioactivity assay. While from the UV220 nm trace, we can clearly observe one large peak that was eluted earlier than the one containing the target mass and its molecular weight correlates to a C-terminal truncate of the precursor peptide which was confirmed by trypsin digestion (Figure 4.6). Since this truncate can also be observed when the LanA was expressed alone, such digestion should not be related to the cyclization and performed by endogenous peptidases in *E. coli*. We also prepared three LanA mutants with the residues next to the putative cleavage site changed to alanine, however, the C-terminal truncates can still be observed, which supports an assumption that the peptide is processed by a C-terminal endogenous peptidase in *E. coli* (Figure 4.7). Therefore, we switched to use an *in vitro* method to get more modified peptides as well as to study the corresponding biosynthetic mechanism.

4.2.4 Optimization of Peptide and Protein Expression and Purification

Originally the precursor peptide was expressed in LB and induced for expression by three hrs, which was a general protocol for preparing LanA. However, due to the endogenous peptidase digestion, the productivity was only 10 μg/L. Therefore, we assume that the peptide productivity can be increased if we shorten the induction time while using a richer medium to maintain the final biomass. Therefore, the expression time was shortened to one hour and the medium was changed to TB, which significantly increased the productivity by ~25 fold (Figure 4.8).

Similar to LanA, a construct for N-terminal his-tagged LanL for expression in *E. coli* was also built. However, through analysis by SDS-PAGE and Western Blot, there was no detectable band with the expected size. Therefore, the LanL was fused with an N–terminal maltose binding protein (MBP) in order to increase the solubility and stability, and a TEV protease digestion site was added between the MBP and LanL, which led to multiple detectable bands on Western Blot (Figure 4.9). While a faint band appeared at the expected range, most of the prominent bands had a smaller size than expected, which supports the fact that the LanL was truncated at its C-terminal. While we tried to remove the MBP-tag by treating the protein eluates by the TEV protease, we observed much insoluble proteins occurred after overnight digestion (Figure 4.9). Through analysis on SDS–PAGE and MALDI–MS, we confirmed that the insoluble fraction should be the LanL, which suggests that the LanL alone has a poor solubility. The instability and possible poor solubility of LanL in *E. coli* may also explain why the productivity of the modified peptide is so low, as there might not be enough LanL to process LanA and the unmodified LanA is also prone to being digested in *E. coli*.

However, in order to study the biosynthetic mechanism of LanL *in vitro*, the LanL should be prepared at a much higher purity. As the lyase and kinase domains are located at the N-terminal half, the C-terminal truncates might still confer partial activity and therefore disturb the reaction stoichiometry *in vitro*. Gel filtration was tried against the eluate from IMAC, which did not significantly increase the purity (Figure 4.10). Considering that the major contaminants are C-terminal truncates of this protein, the N-terminal his-tag was therefore moved to the C-terminus of this fusion protein to avoid the coelution of truncates together with the full length protein, which not only yields the protein with much higher purity but also simplifies the protein purification process (Figure 4.10).

4.2.5 *In vitro* Characterization of the Biosynthesis of the Class IV Lanthipeptide

After optimization of the peptide and protein preparation method, we successfully reconstituted the reaction *in vitro* (Figure 4.11). Compared with the unmodified precursor peptide which was modified by two NEM, only 2-fold dehydration was observed for the modified peptide, indicating both dehydration and cyclization reactions were completed *in vitro*. By scaling up the reaction by using 3 mg precursor peptide as substrate, we successfully got sufficient amount of modified peptide for determination of the stereochemistry of thioether crosslinks (Figure 4.12). After hydrolysis, derivatization and GC/MS analysis, it appears that both thioether crosslinks in our lanthipeptide are in the form of DL-methyllanthionine.

To further study the biosynthetic mechanism, the *in vitro* reaction time was shortened to 10 min in order to get the biosynthetic intermediates. Fortunately, a series of intermediates were observed on MALDI-TOF-MS. After digestion by trypsin, the corresponding fragments were analyzed on

both MALDI-TOF-MS as well as MS/MS in order to get more structural information (Figure 4.11 and Figure 4.14). The structures of all intermediates detected reveal that there might be multiple routes for the biosynthesis of this lanthipeptide. Although there are some evidence showing that the dehydration of venezuelin happens at an N-C direction, both dehydration and cyclization of this lanthipeptide can happen from both N-C and C-N direction.

To further validate this hypothesis as well as to explore if there is any biosynthetic route that the lanthipeptide synthase prefers, two mutants of the precursor peptides were prepared (T49A and T52A) for *in vitro* reaction (Figure 4.13 and Figure 4.14). After 1h reaction, T49A was completely converted to the corresponding product with the C–terminal ring formed. However, T52A was partially converted into the product with the N–terminal ring formed while the intermediate with 1-fold dehydration was also observed, which means that the cyclization may prefer the C–N direction. Of note, another intermediate with 2-fold dehydration was also observed, which implies that the lyase/kinase domain(s) might not be very specific.

4.3 Conclusion

In this chapter, we successfully applied our pathway refactoring workflow to the discovery of a novel Class IV lanthipeptide from *Streptomyces sp*. NRRL S-1022. While very limited members of Class IV lanthipeptides have been discovered to date and all of them exhibit similar structural features (e.g. venezuelin and its homologs), the lanthipeptide discussed in this chapter exhibits a completely novel ring topology. To further study its biosynthetic mechanism, the expression and preparation process for both the precursor peptide and the synthase was optimized. The biosynthetic reaction for this lanthipeptide was successfully reconstituted *in vitro* and the analysis

of its intermediates reveal its biosynthetic route. Multiple biosynthetic routes should co-exist during the biosynthesis. While the dehydration reactions may not have strong preference from either N–C or C–N direction compared with the biosynthesis of venezuelin which confers an N–C dehydration order, the cyclization reactions should prefer a C–N direction, although it can also happen in both ways. Overall, in this chapter we not only discovered a novel Class IV lanthipeptide with unprecedented structural features, but also generate more knowledge about the biosynthesis of Class IV lanthipeptides.

4.4 Materials and Methods

4.4.1 Bacterial Strains and Materials

E. coli DH5a cells were used for cloning and mutagenesis. *E. coli* BL21(DE3) cells were used for expression. The *Streptomyces sp*. NRRL S-1022 strain used in this study was obtained from the United States Department of Agriculture Agricultural Research Service (USDA ARS) collection (Peoria, IL). All plasmids were validated via sequencing (carried out by ACGT, Inc.). All the enzymes for cloning were obtained from New England Biolabs (Ipswich, MA), oligonucleotide primers from the Integrated DNA Technologies (Coralville, IA), and trypsin from Worthington Biochemical Corp (Lakewood, NJ). Lysozyme, benzonase (25–29 U/μL), Millipore C18 Ziptips were purchased from Thermo Fisher Scientific (Waltham, MA).

4.4.2 Coexpression of the Precursor Peptide and Synthase and Purification of the Modified Peptide

The pET28a plasmid containing the precursor peptide gene and the synthase was constructed by the plug-and-play workflow as described previously (23). Two 4 L baffled flasks, carrying 2 L of

LB each, were inoculated 1:100 with an LB overnight culture and grown at 37 °C until reaching an OD600 of 0.6–0.8. Cultures were induced by addition of 1 mL of IPTG stock solution (1 M) per flask and continued shaking under 18 °C for another 18 – 20 h. Cells were harvested by centrifugation and resuspended in LanA start buffer (20 mM NaH_2PO_4, pH 7.5 at 25 °C, 500 mM NaCl, 0.5 mM imidazole, 20% glycerol) and lysed by sonication with the following settings: 35% amplitude, 15 min total sonication time; 4.0 s on / 9.9 s off-pulse. The cell lysate was then centrifuged at 18,000 rpm for 30 min and the supernatant was collected as the soluble fraction. While the precipitant was dissolved into the guanidine lysis buffer (6 M guanidine hydrochloride, 20 mM NaH_2PO_4, 500 mM NaCl, 0.5 mM imidazole, pH 7.5), sonicated and centrifuged again by using the same setup, and the supernatant was collected as the insoluble fraction.

Either the soluble fraction or insoluble fraction was then purified by IMAC. The 5 mL HisTrap column (GE Healthcare) was equilibrated by guanidine lysis buffer and the sample was then loaded by a peristaltic pump. The HisTrap column was then washed by guanidine wash buffer (4 M guanidine hydrochloride, 20 mM NaH_2PO_4, 500 mM NaCl, 30 mM imidazole, pH 7.5) and eluted with 15 mL of guanidine elution buffer (4 M guanidine hydrochloride, 20 mM Tris, 100 mM NaCl, 1 M imidazole, pH 7.5). The elution fraction was desalted by BondElute C18 solid phase extraction column (Agilent). After lyophilization, the peptide powder was dissolved into 5 mL water with 0.1% (v/v) trifluoroacetic acid (TFA) and cleaned by 0.22 um filter for semiprep HPLC purification with 0.1% (v/v) TFA (A) and MeCN–0.1% (v/v) TFA (B) at 3 mL/min and the following gradient: linear gradient of 5%–80% B over 30 min, followed by a linear increase of 80–100% B over 1 min and holding 100% B for another 5 min. The elution fractions containing the target peptide were lyophilized and re-dissolved in 0.1% (v/v) TFA and stored under -80 °C.

4.4.3 Expression and Purification of the Precursor Peptide

The pET28a plasmid containing the precursor peptide gene under control of T7 promoter was expressed in *E. coli* BL21(DE3). Two 4 L baffled flasks, carrying 2 L of TB each, were inoculated 1:100 with an LB overnight culture and grown at 37 °C until reaching an OD600 of 0.6–0.8. Cultures were induced by addition of 200 μL of IPTG stock solution (1 M) per flask and harvested after expression for 1 h at 37 °C. Cells were resuspended in guanidine lysis buffer (6 M guanidine hydrochloride, 20 mM NaH_2PO_4, 500 mM NaCl, 0.5 mM imidazole, pH 7.5) to a final volume of 35–40 mL and lysed at room temperature by sonication with the following instrument settings: 60% amplitude, 5 min total sonication time; 5 s on / 5 s off-pulse.

After centrifugation for 30 min at 18,000 rpm and 4 °C, the supernatant was sonicated again for 1 min at the same settings to further shear the DNA to allow easy filtration through a syringe filter (0.45 μM). After filtering, the lysates were applied to IMAC purification as described above. In this way, the precursor peptide could be obtained with yields of ~0.25 mg/L expression culture. The elution fractions containing the target peptide were lyophilized, redissolved in 0.1% (v/v) TFA with the concentration determined by NanoDrop (ThermoFisher, Waltham, MA) and stored under -80 °C.

4.4.4 Protein Expression and Purification

The MBP tagged LanL was expressed for 18 – 20 h at 18 °C in 4 L baffled flasks containing 2 L of TB medium using the conditions described above. Cell pellets from 4 L of cultures were resuspended in lysis buffer, yielding a final volume of 35–40 mL. Cells were incubated on ice with

lysozyme (50 mg) and benzonase (5 μL) for 1 h and subsequently lysed by sonication on ice with the following settings: 60% amplitude, 5 min total sonication time; alternating between 2 s on / 5 s off-pulse.

Lysates were centrifuged twice for 30 min at 4 °C and 18,000 rpm (supernatants were transferred into fresh centrifuge tubes in between centrifugation steps) and filtered through a syringe filter (0.45 μm). IMAC purification was performed using a 5 mL HisTrap column (GE Healthcare) using fast protein liquid chromatography (FPLC) with buffer A (300 mM NaCl, 50 mM NaH_2PO_4, 20 mM imidazole, pH 8.0) and buffer B (300 mM NaCl, 50 mM NaH_2PO_4, 250 mM imidazole, pH 8.0) at 5 mL/min. After loading, the column was washed by approximately 100 mL buffer A till the UV 280 nm absorption fell below 20 mAU and then eluted with the following gradient: linear gradient of 0 – 100% B over 9 min followed by 100 % B for another 6 min. Elution fractions were cleaned by 0.22 μm filter and concentrated with Amicon centrifugal filter units with 100 kDa MW cut-off to around 500 μL. The original buffer was also exchanged to the storage buffer (300 mM NaCl, 50 mM HEPES, 10% glycerol at pH7.5) and the concentration was determined by NanoDrop. The protein was then snap – freezed and stored at -80 °C.

4.4.5 *In vitro* Enzymatic Assays

A typical enzyme assay contained 20 μL of 2× buffer (300 mM NaCl, 50 mM HEPES, pH 7.5), 2 μL of tris(2-carboxyethyl)phosphine (TCEP) stock solution (20 mM in H_2O), 2 μL of $MgCl_2$ stock solution (200 mM in H_2O), 2 μL of ATP stock solution (50 mM in H_2O), 2 μL of the precursor peptide stock solution (10 mg/mL in 0.1% (v/v) TFA, which yields a final concentration of 65 μM in the assay) and corresponding amounts of protein stock solution and H_2O to obtain a final volume

of 40 μL at a concentration of 2.5 μM for the protein. The reaction was run overnight at room temperature and then 20 μL of the sample was purified and eluted into 10 μL of 70% (v/v) MeCN/0.1% (v/v) TFA using C18 Ziptips following the manufacturer's protocol. 1 μL samples were prepared for MALDI-TOF-MS by mixing with 1 μL of 15 mg/mL α-cyano-4-hydroxycinnamic acid (CHCA) in 70% (v/v) MeCN/0.1% (v/v) TFA.

4.4.6 Determination of the Reaction Order

To determine the order of the dehydration and cyclization, 60 μL assay reactions were set up as described above and the reaction was either stopped after 10 min or 60 min. 20 ul aliquots of the reaction were quenched in a thermocycler by holding the temperature at 95 °C for 5 min. The reaction mixture was then lyophilized, redissolved by 20 μl guanidine buffer with 1 μL of TCEP stock solution (100 mM in H_2O) and incubated for 15 min at 50 °C. Then, 1 μL of NEM stock solution (400 mM in EtOH) was added, and after mixing, the reaction was run for 1 h at room temperature. The peptide was purified and eluted into 20 μL of 70% (v/v) MeCN/0.1% (v/v) TFA using C18 Ziptips following the manufacturer's protocol and prepared for MALDI-TOF-MS by mixing 1 μL of sample with 1 μL of 15 mg/mL CHCA in 70% (v/v) MeCN/0.1% (v/v) TFA to confirm the existence of intermediates. The sample prepared by C18 Ziptip was lyophilized again and redissolved into 50 mM Tris HCl, 1 mM $CaCl_2$, pH 7.6 with 0.15 mg/mL trypsin. After incubation for 3 h at room temperature, the reaction was purified and eluted into 20 μL of 70% (v/v) MeCN/0.1% (v/v) TFA using C18 Ziptips again and analyzed by MALDI-TOF-MS. For MS/MS analysis on Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Scientific), the sample was lyophilized again and redissolved into 0.1% (v/v) formic acid.

4.5 Figures

Figure 4.1. Alignment of the putative precursor peptide sequences from GCF128.

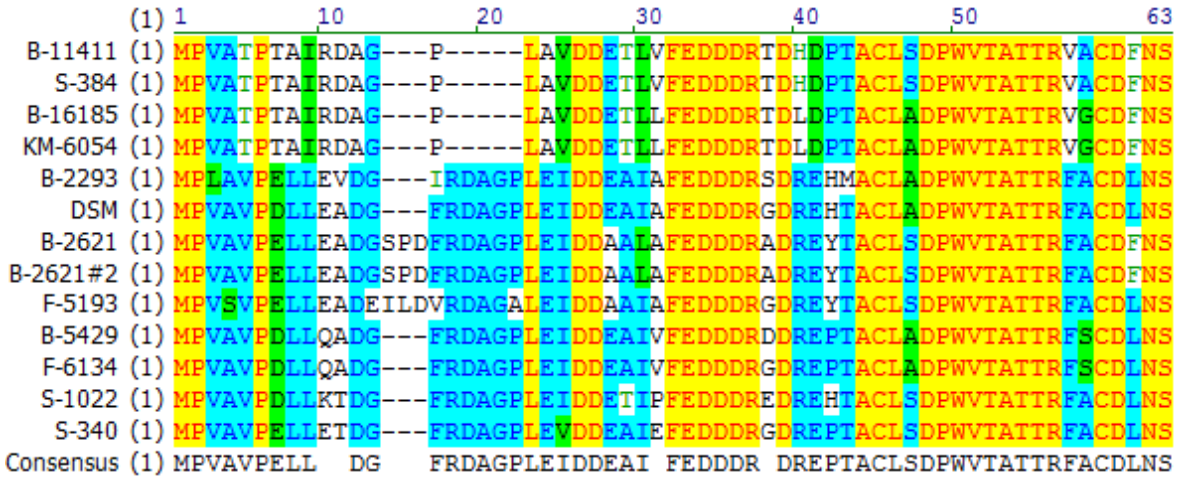


Figure 4.2. Modified precursor peptide with 2-fold dehydration (m/z observed: 7714.4) and unmodified precursor peptide modified by two NEM (m/z observed: 8000.4).

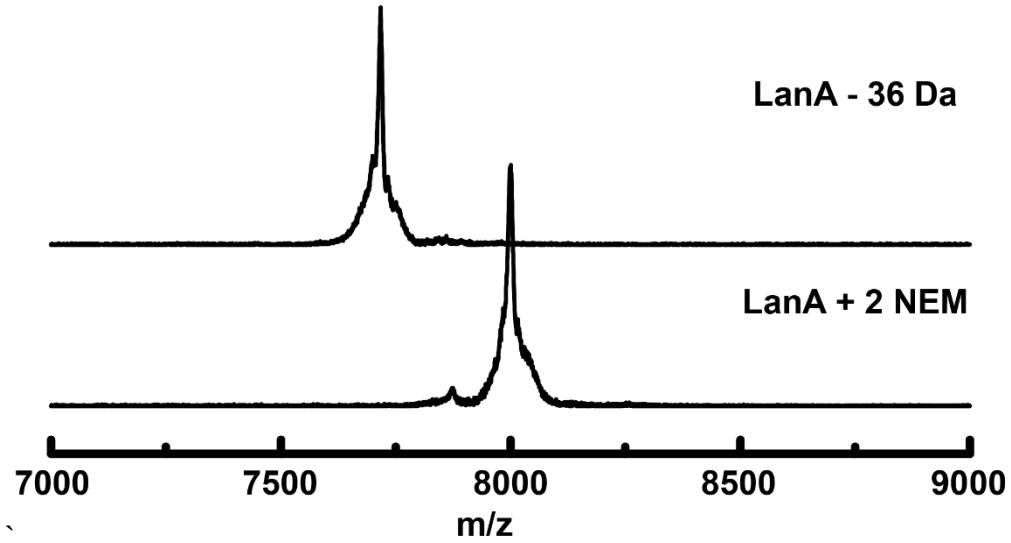


Figure 4.3. The modified peptide digested by trypsin.

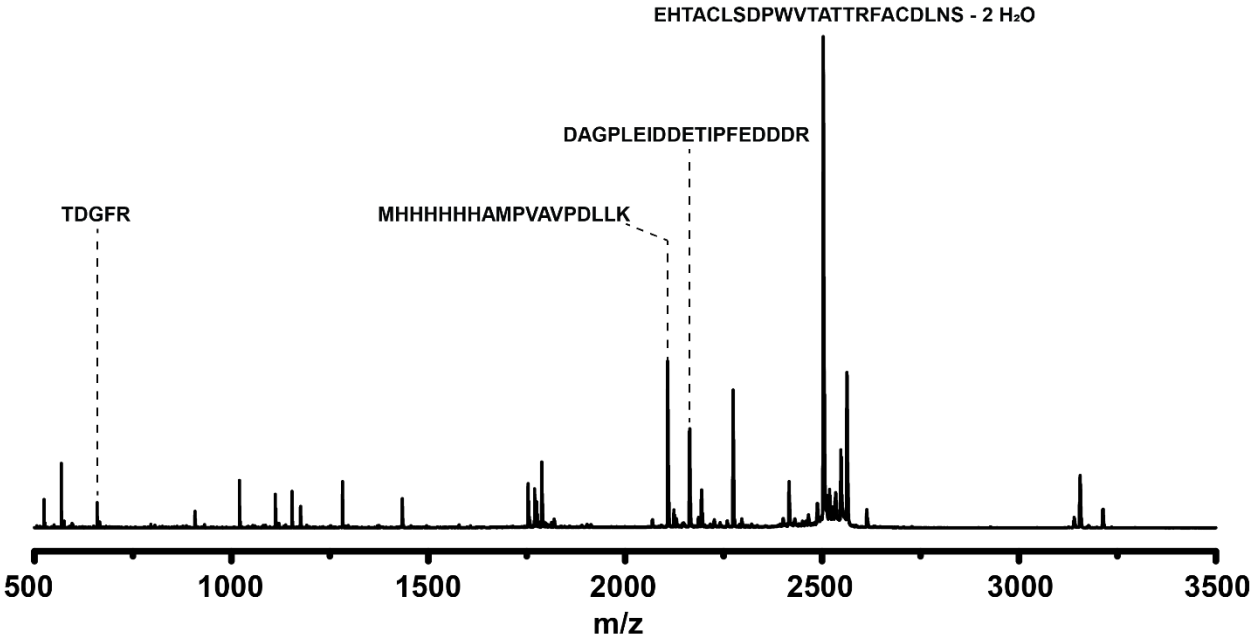


Figure 4.4. MS/MS analysis for the modified peptide.

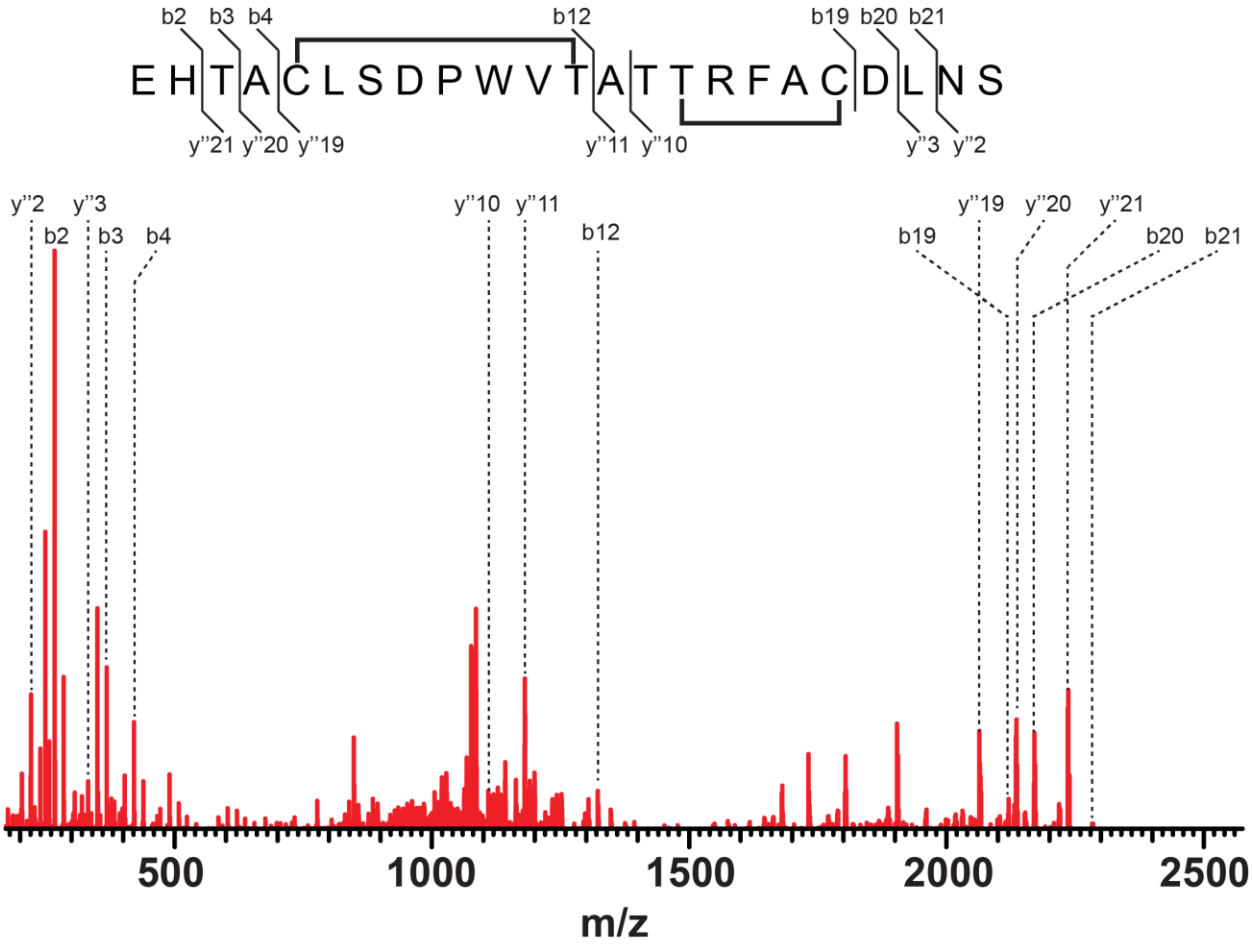


Figure 4.5. Coexpression of the precursor peptide mutants with LanL.

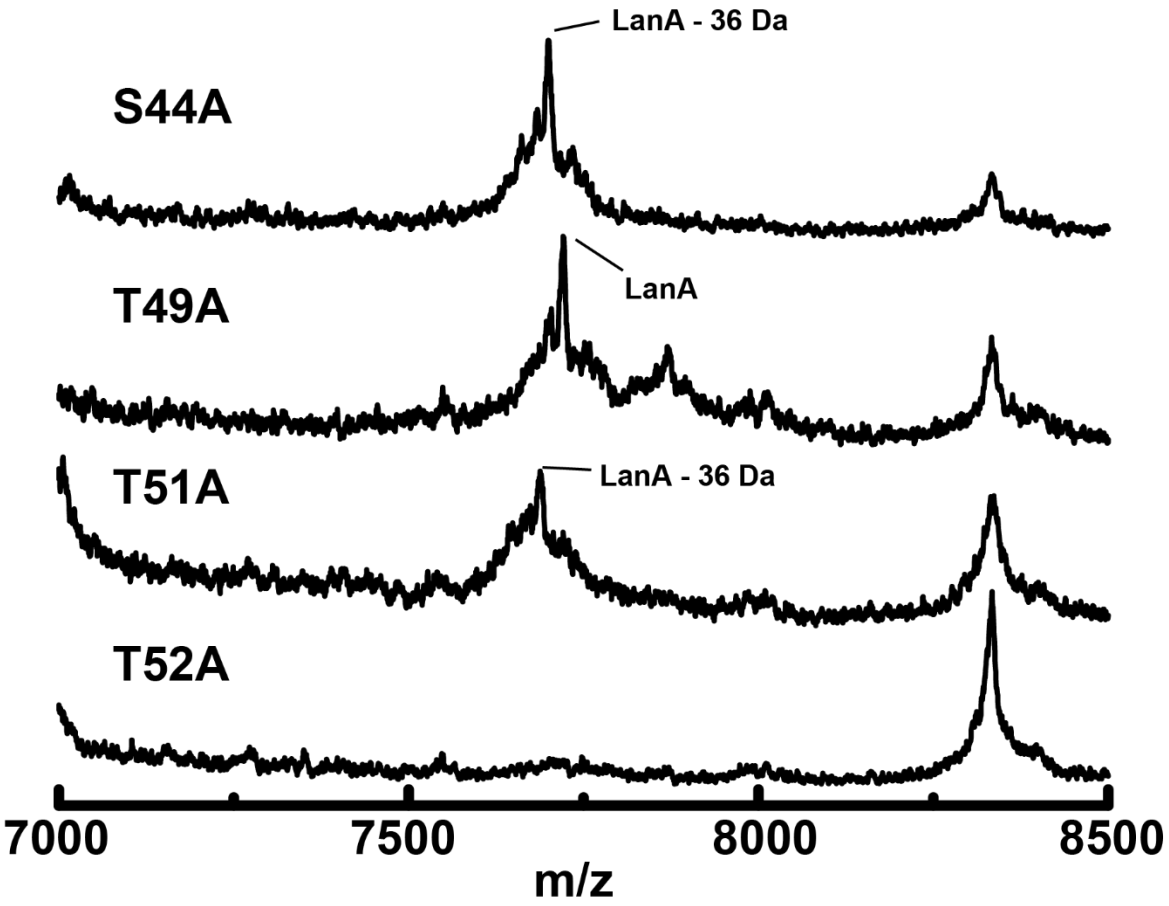
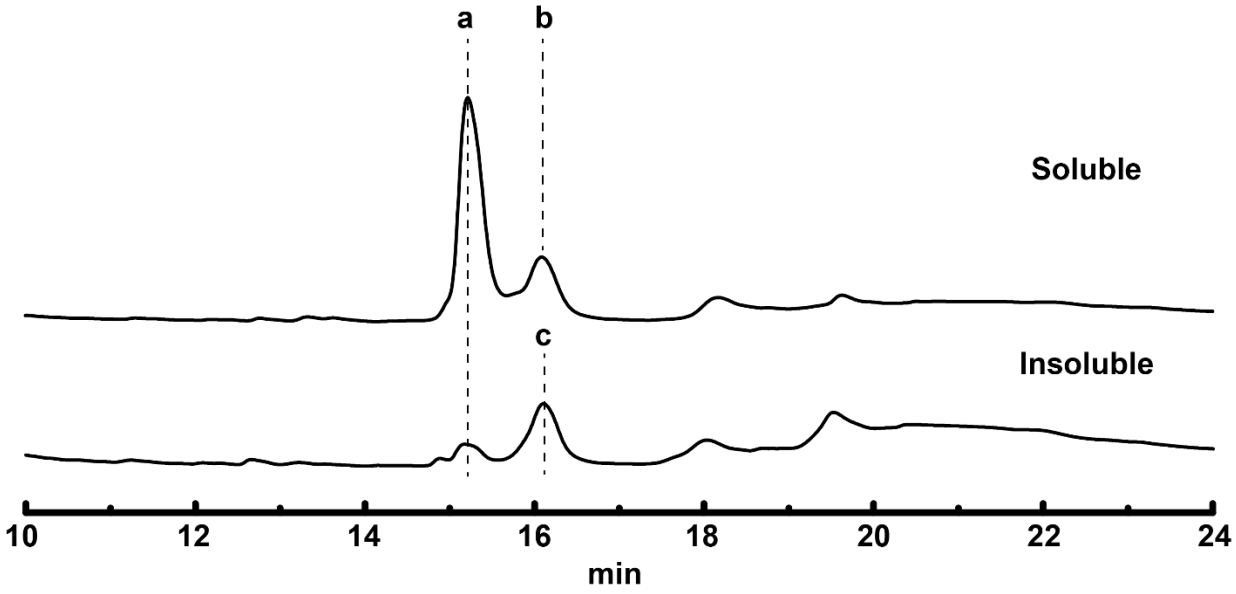


Figure 4.6. A) UV220nm trace for soluble and insoluble fractions. a: C-terminal truncated peptide; b: modified peptide; c: unmodified peptide. B) Fraction a digested by trypsin.



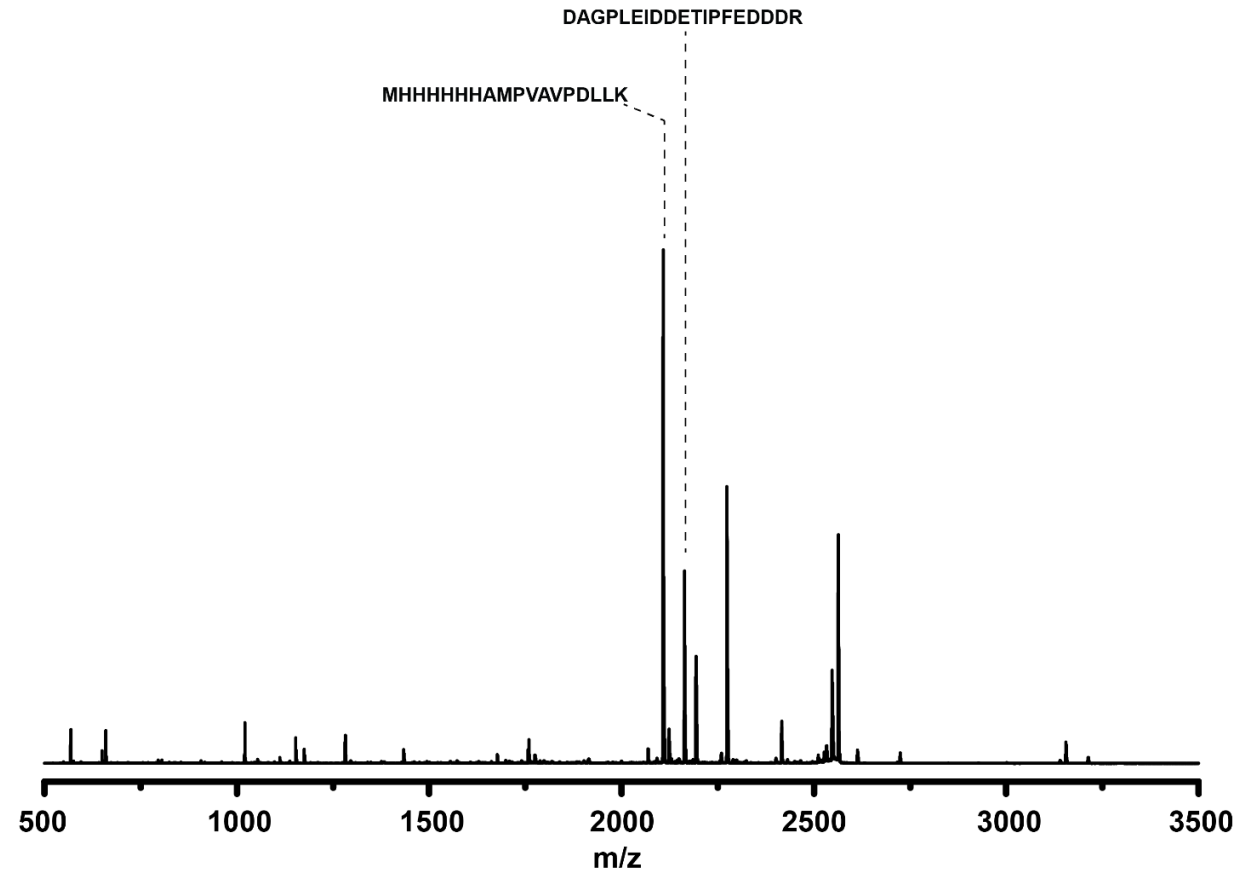


Figure 4.7. MALDI-TOF-MS results for desalted IMAC purified *E. coli* cell lysates with LanA mutants coexpressed with LanL.

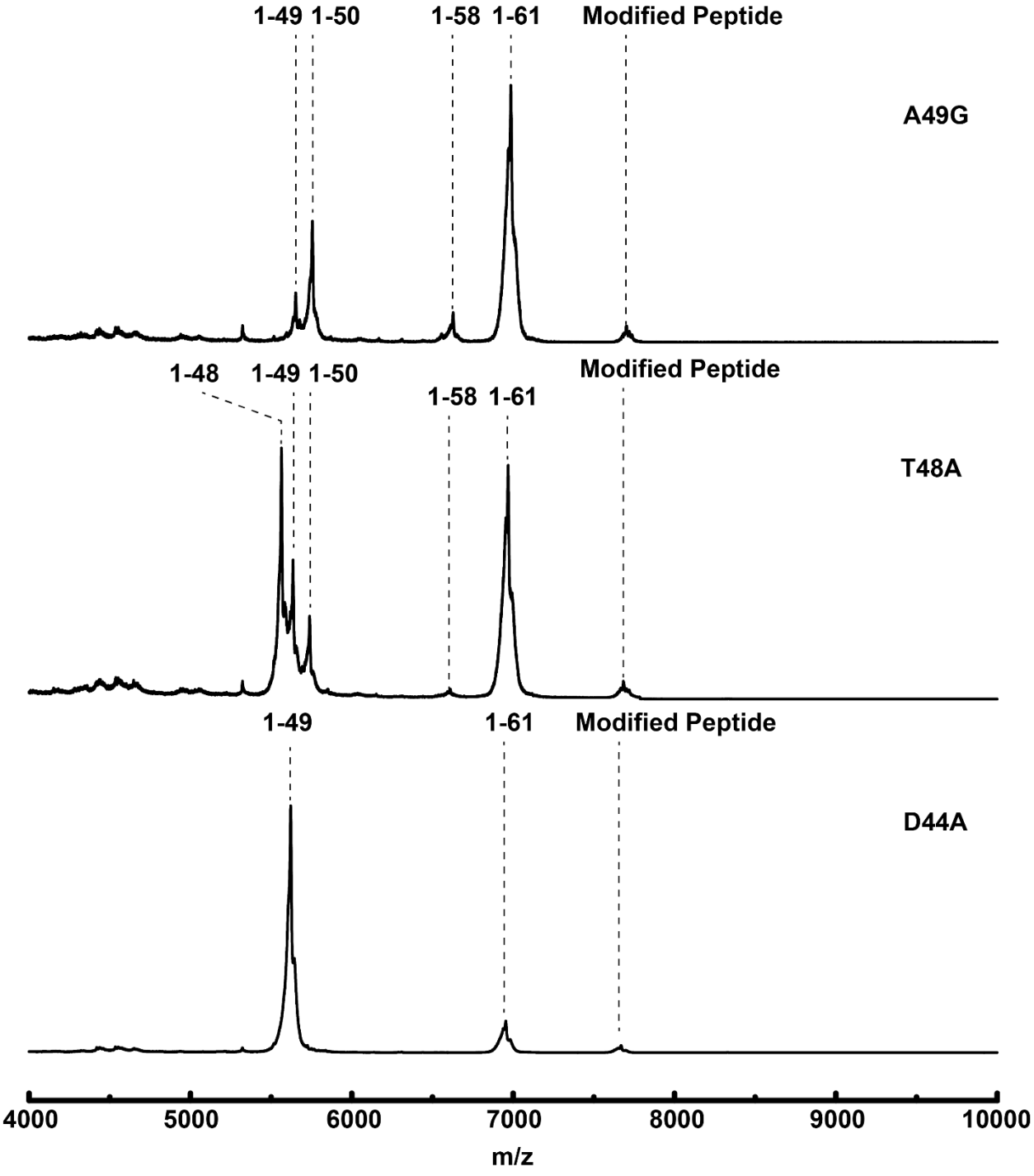


Figure 4.8. UV220nm trace for purification of LanA on semiprep HPLC. Peaks containing the peptide are highlighted.

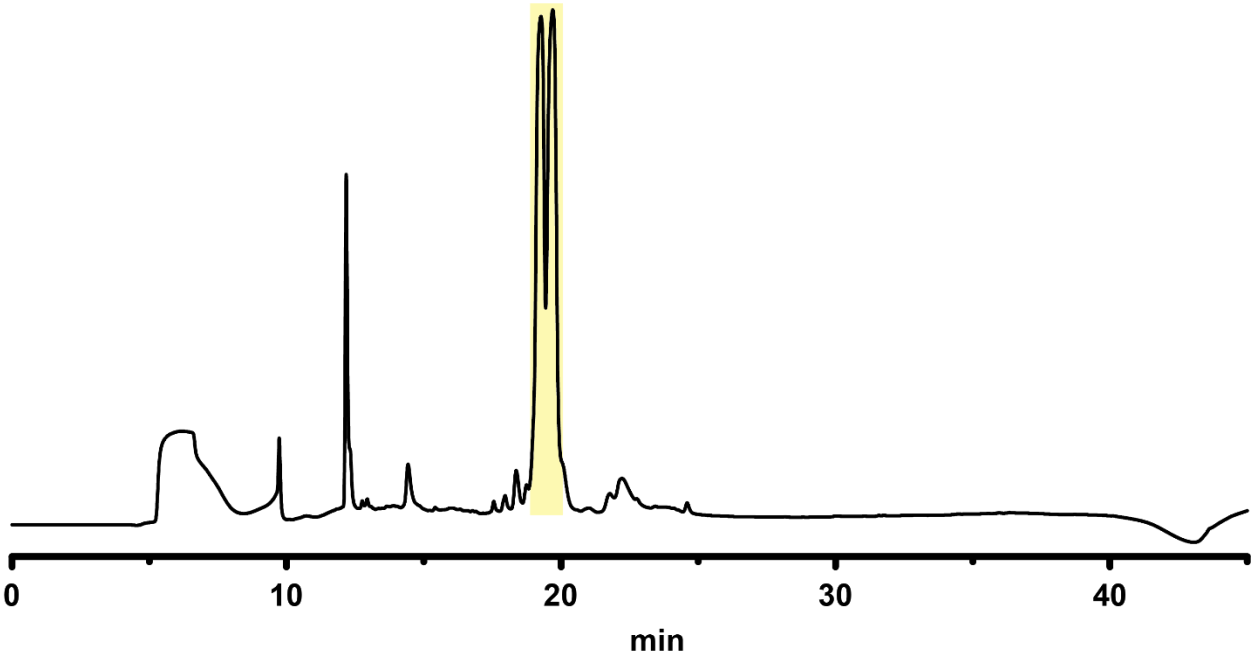


Figure 4.9. A) The organization and the corresponding size of the MBP-LanL fusion protein. B) SDS-PAGE of the fractions from IMAC purification. C) Western Blot of the fractions from IMAC purification. D) SDS-PAGE of the IAMC eluates digested by TEV protease. E) MALDI-TOF-MS analysis of the insoluble fraction after TEV digestion.

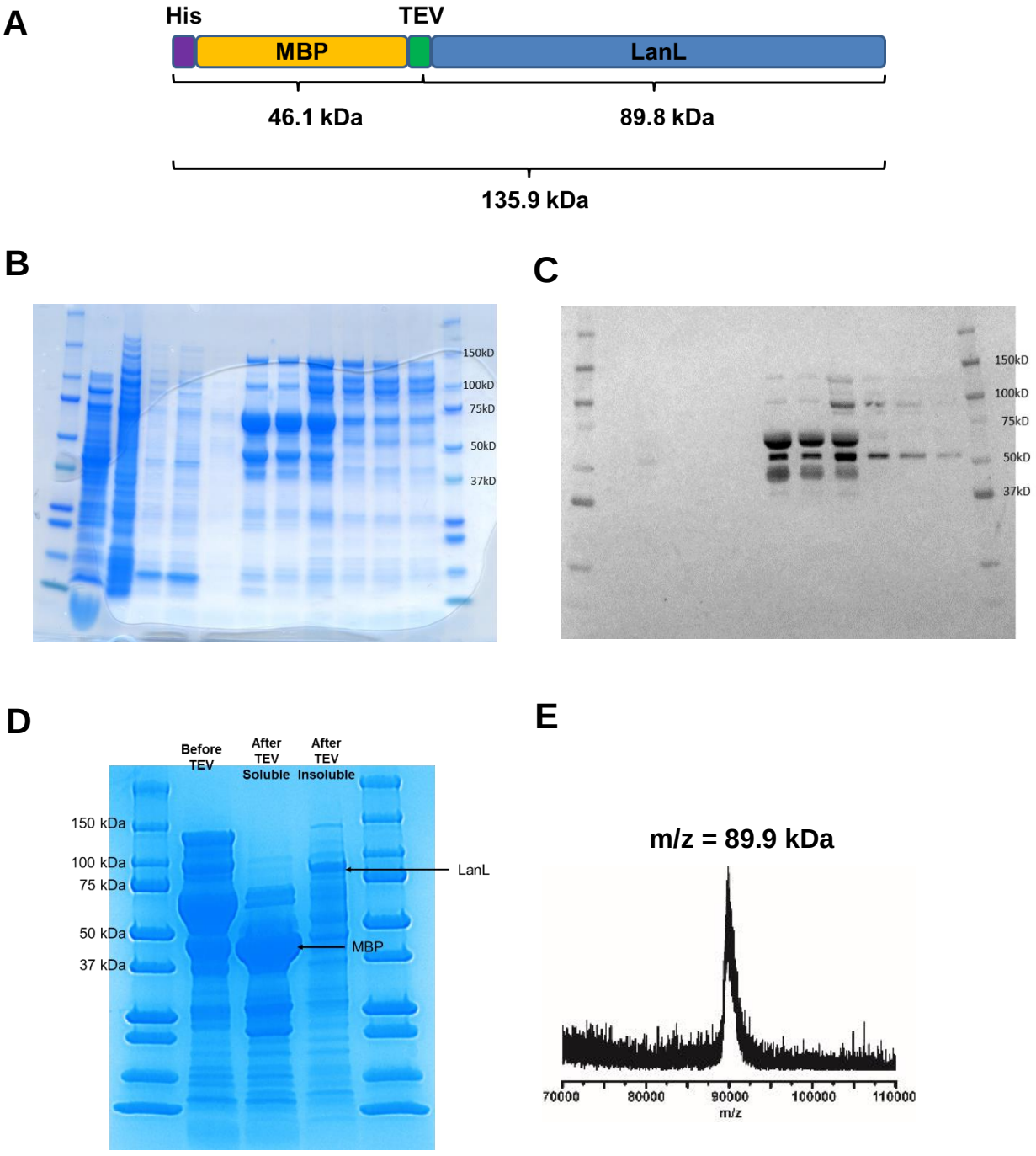


Figure 4.10. A) UV 280nm trace for purification of the MBP-LanL IMAC eluates on HiLoad 16/600 Superdex 200 pg gel filtration column. B) SDS-PAGE of fractions from gel filtration purification. C) SDS-PAGE of IMAC purified MBP-LanL with C-terminal His-tag.

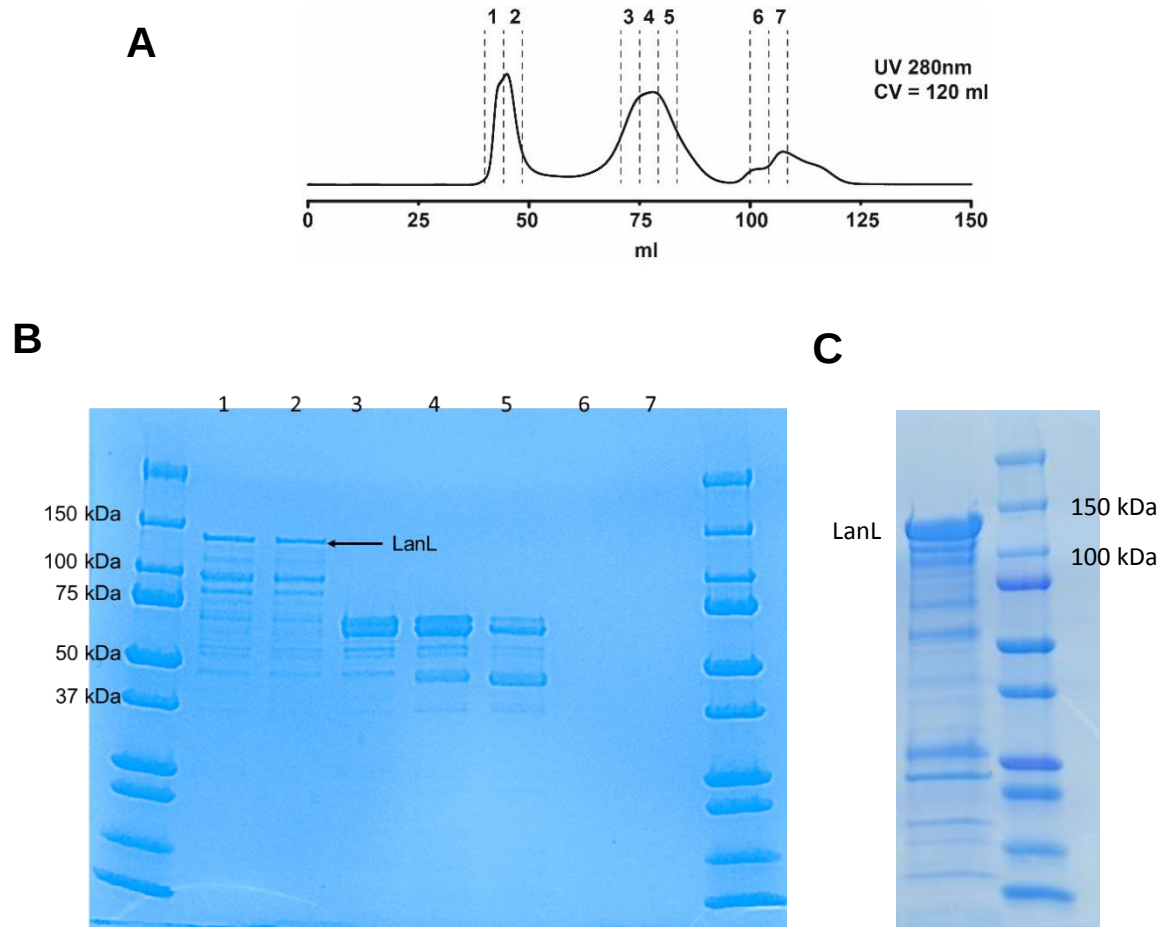


Figure 4.11. A) *In vitro* enzymatic assay stopped at either 10 min or 60 min. B) The trypsin digested results of 10 min reaction on MALDI-TOF-MS. C) The trypsin digested results of 60 min reaction on MALDI-TOF-MS.

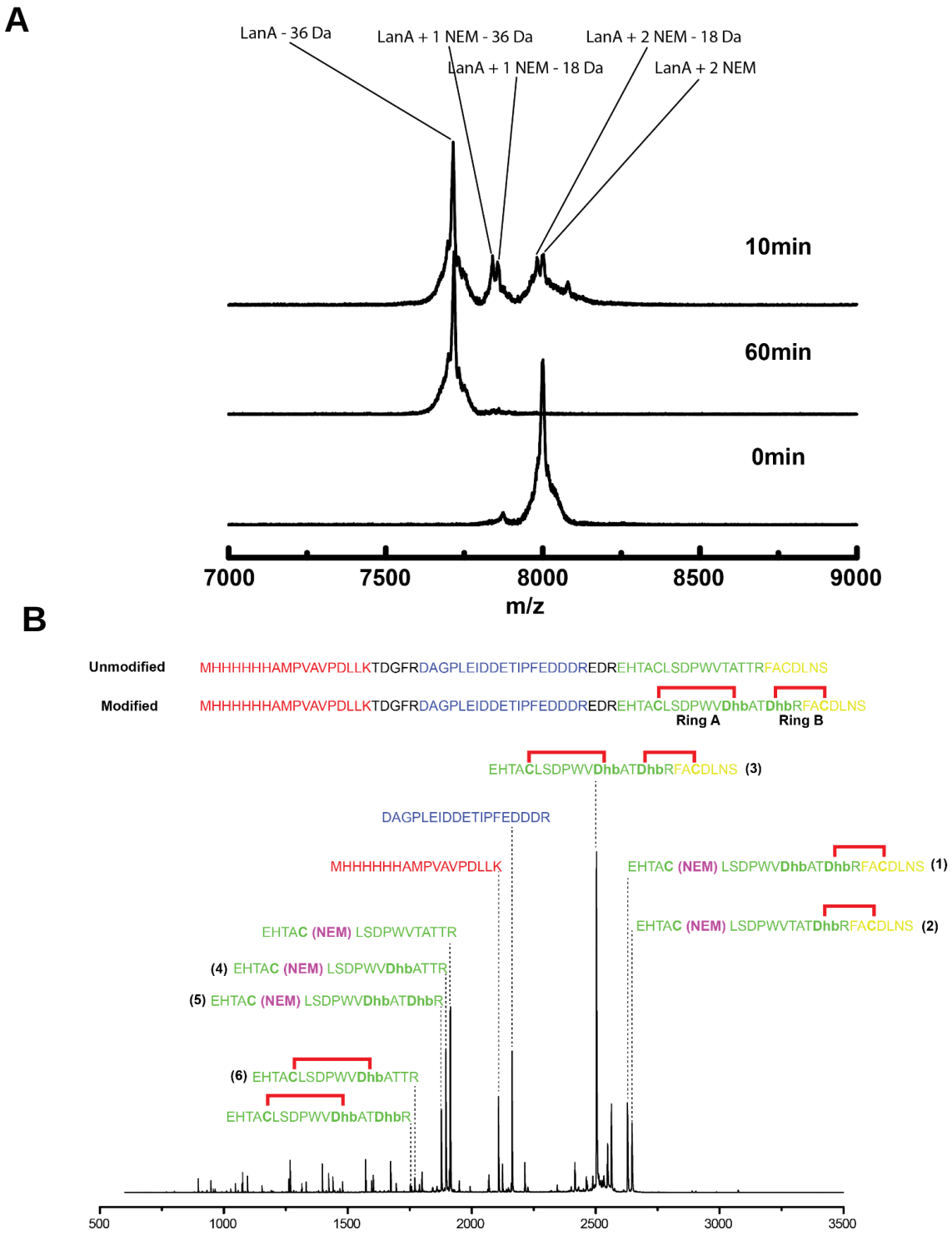


Figure 4.11. Continue

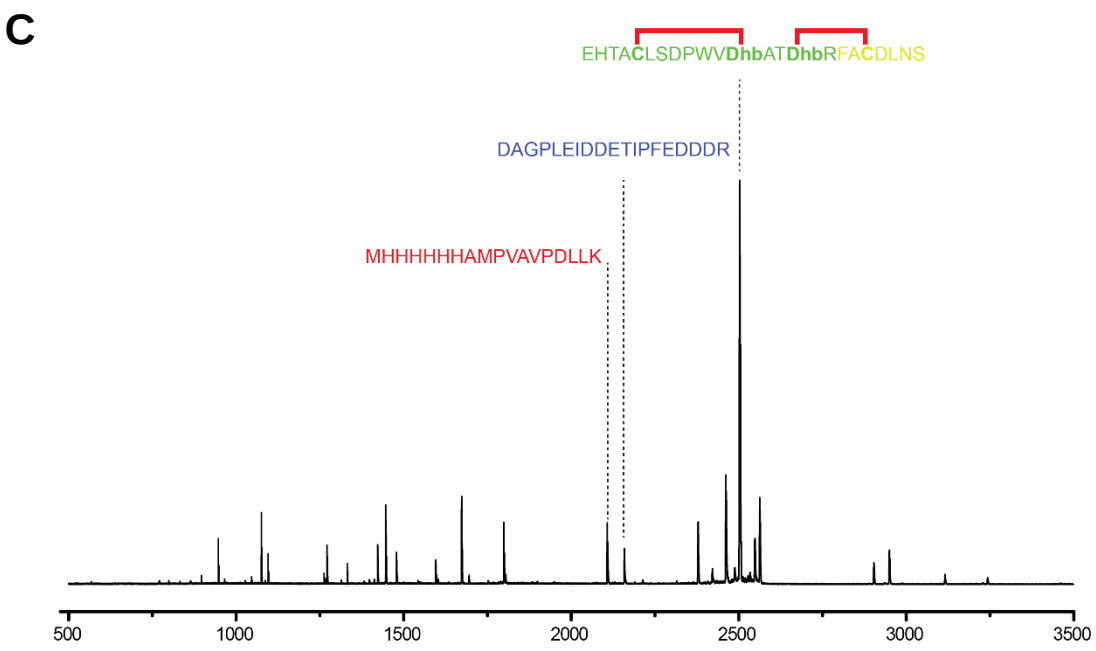


Figure 4.12. Extracted ion chromatogram of the stereochemistry determination of thioether crosslinks on GC/MS. Figure made by Dr. Ian Bothwell.

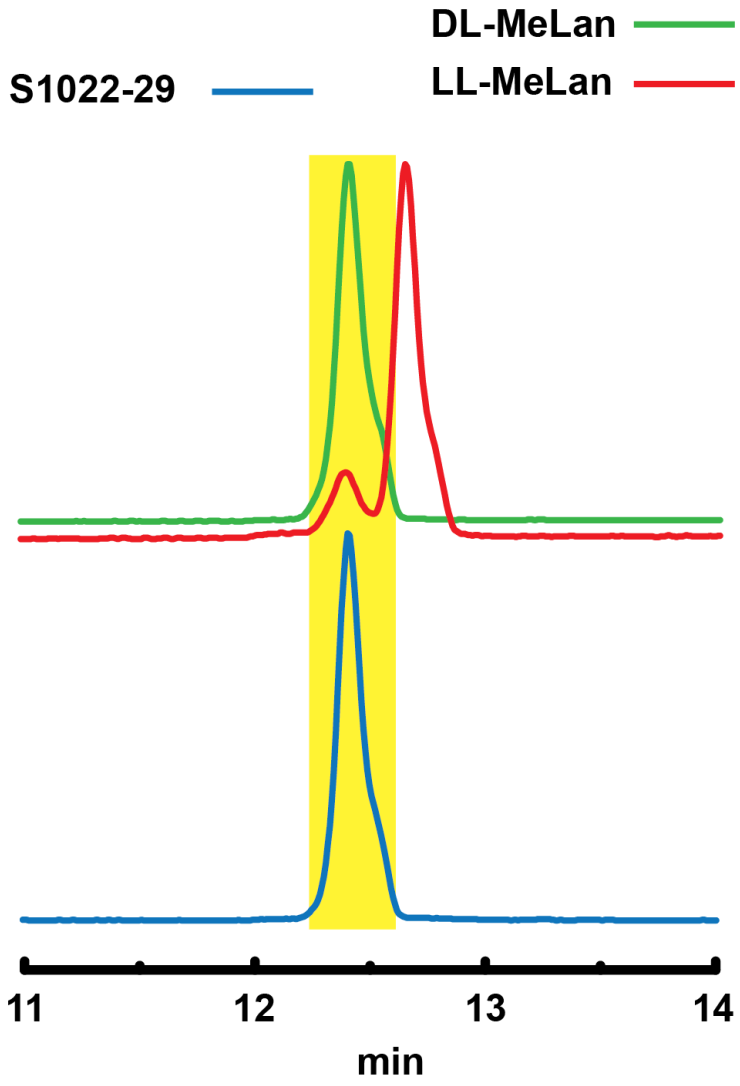


Figure 4.13. *In vitro* enzymatic reaction using T49A and T52A as substrates. Peptide fragments were obtained by trypsin digestion and analyzed on MALDI-TOF-MS.

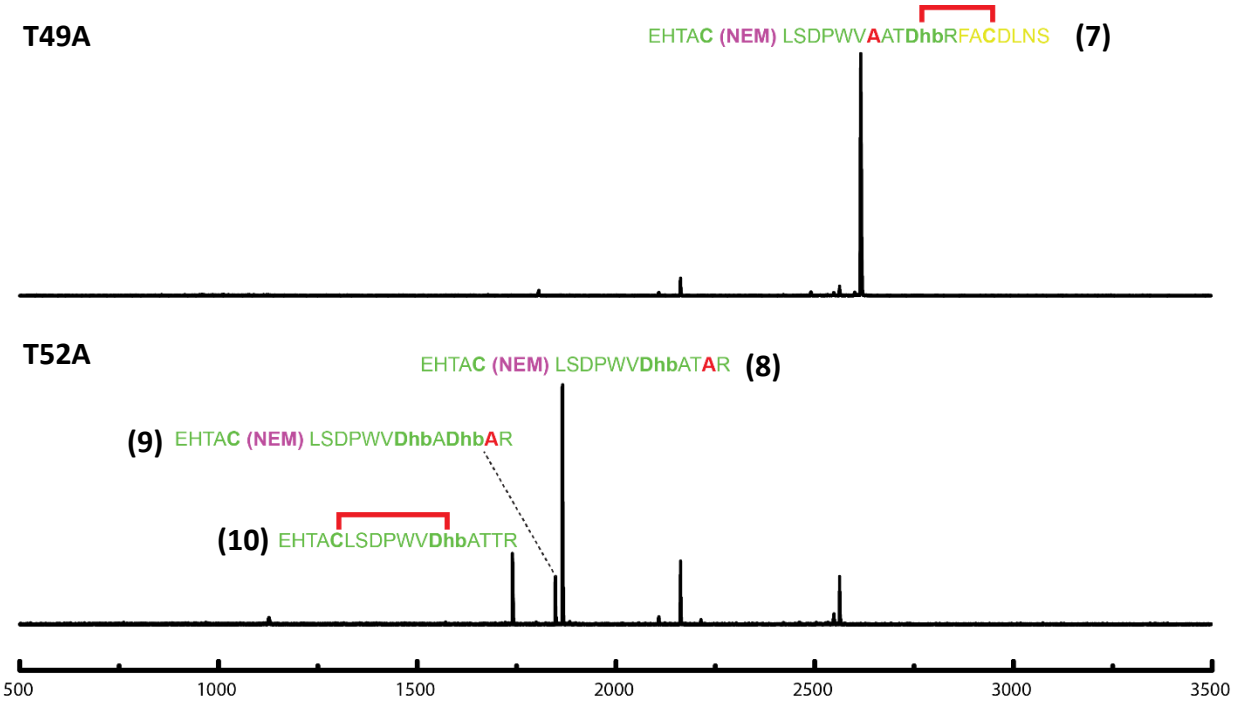
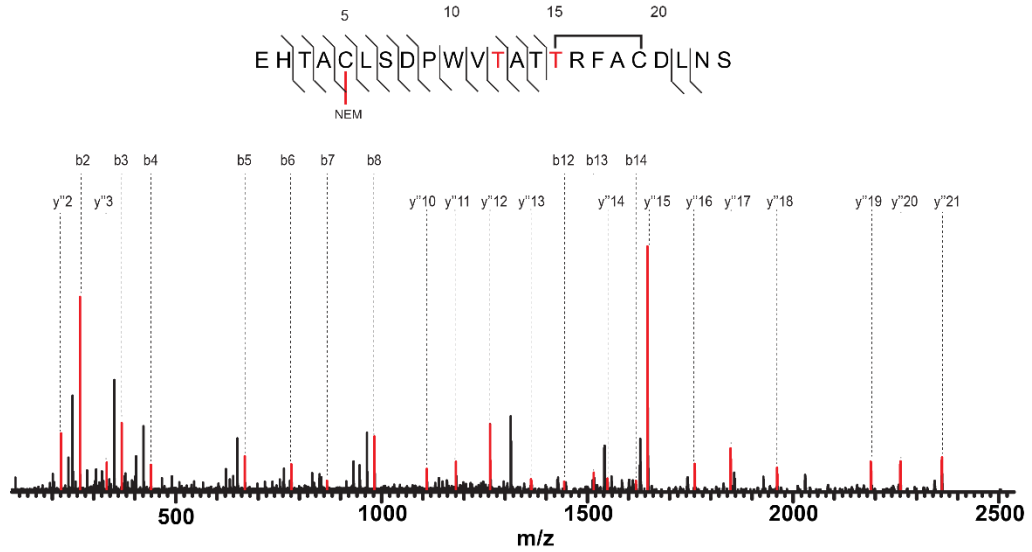


Figure 4.14. MS/MS analysis of all peptide fragments obtained from *in vitro* enzymatic reaction after trypsin digestion.

(1)



(2)

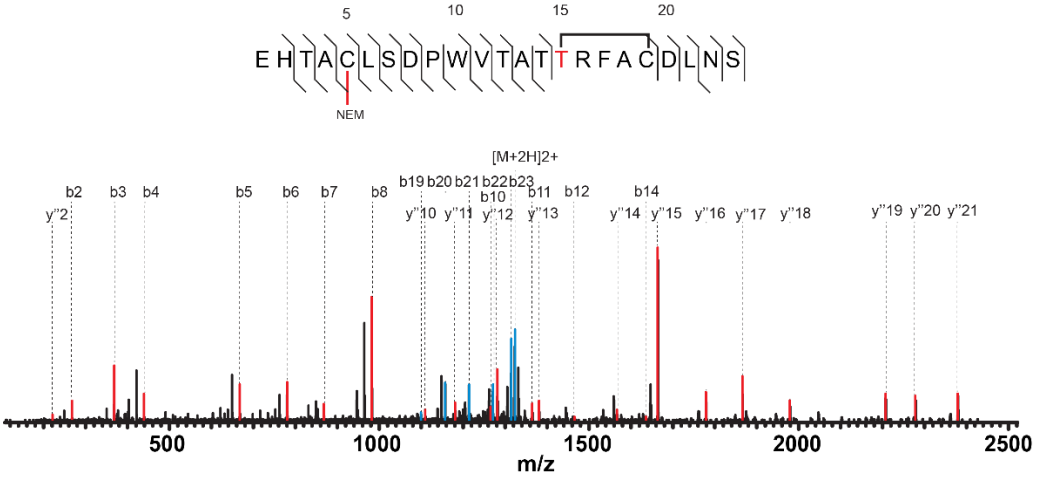
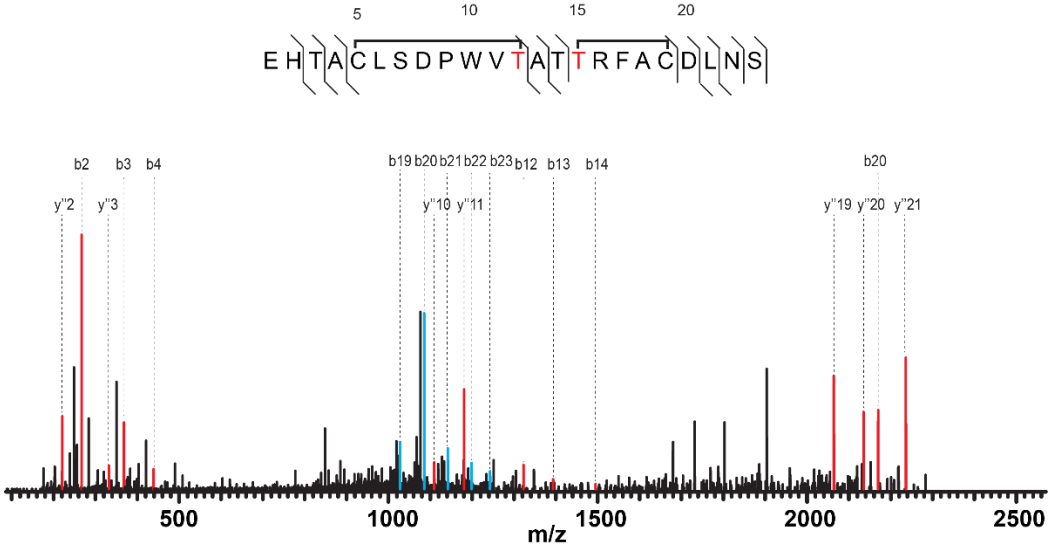


Figure 4.14. Continue

(3)



(4)

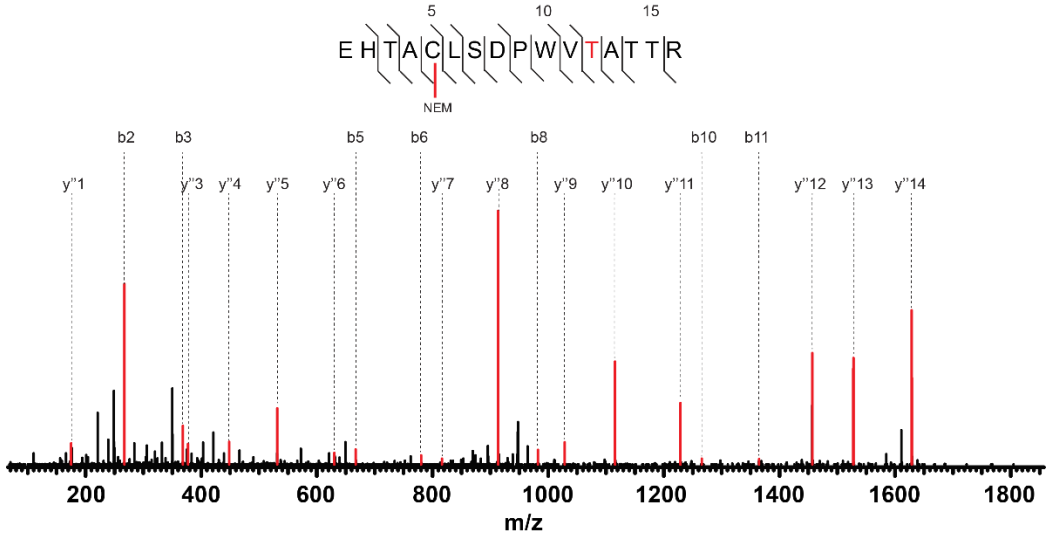
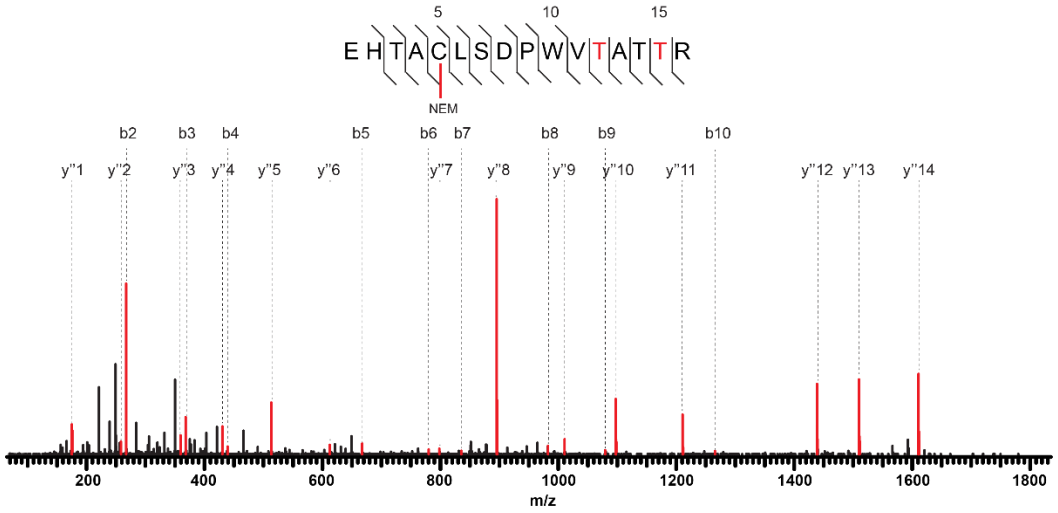
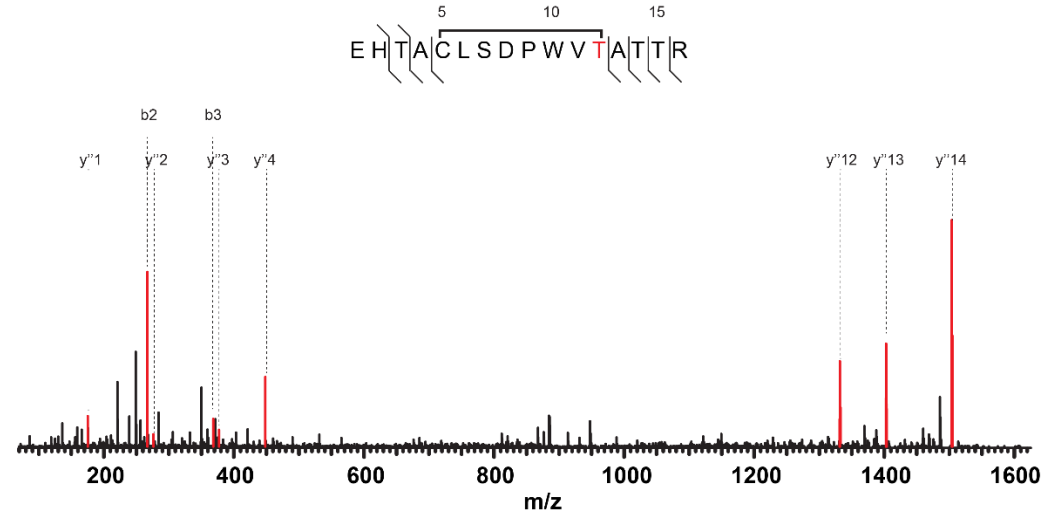


Figure 4.14. Continue

(5)



(6)



(7)

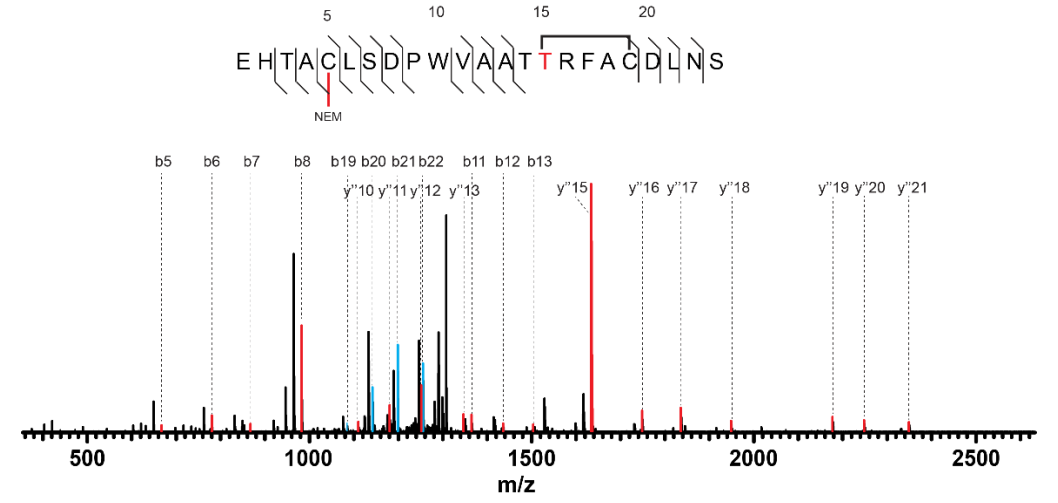
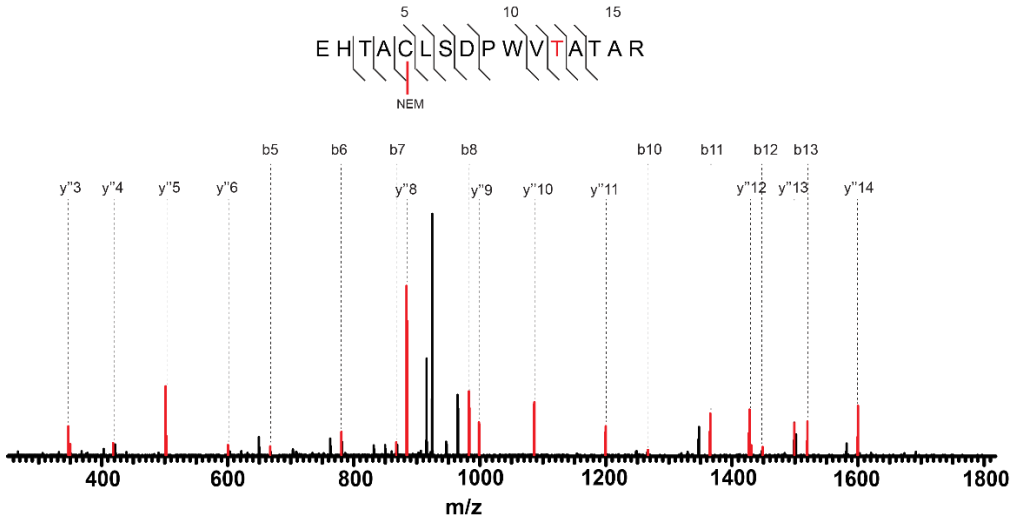


Figure 4.14. Continue

(8)



(9)

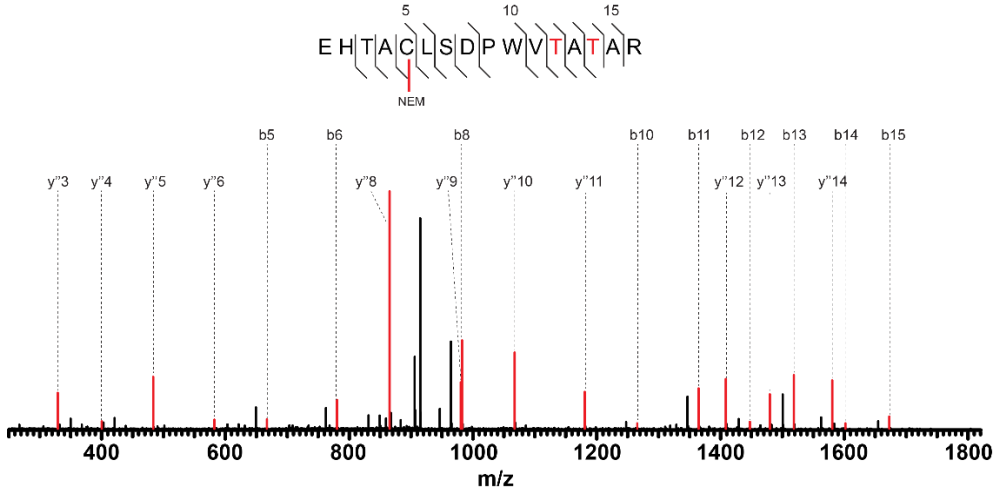
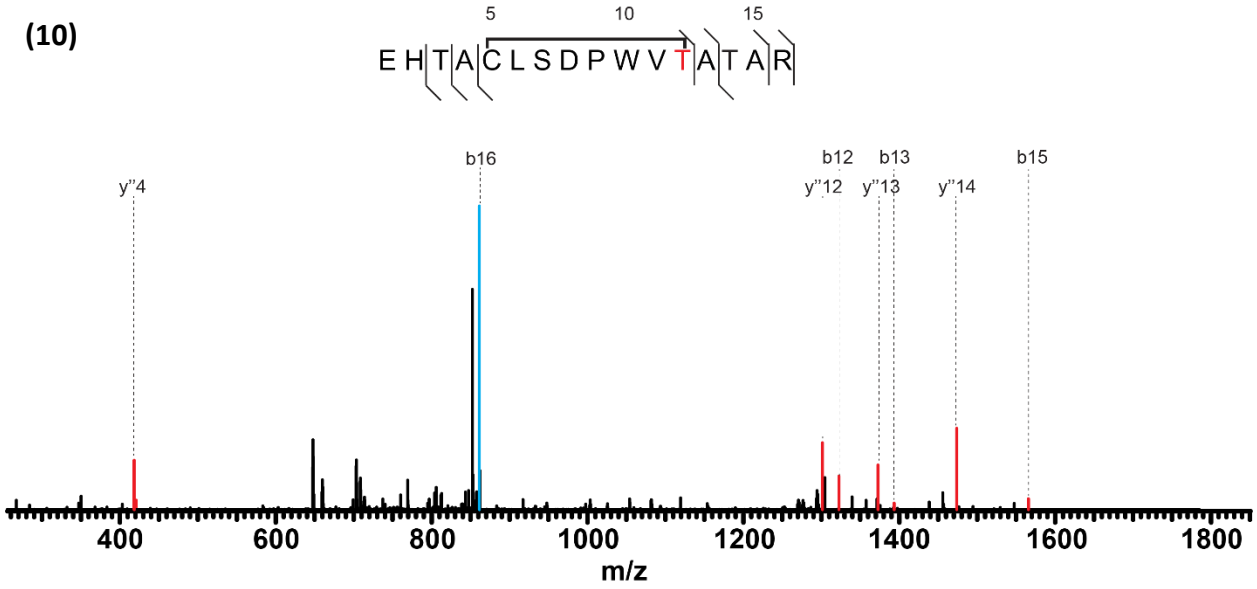


Figure 4.14. Continue

(10)



4.6 References

1. Arnison PG, *et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30(12):1568-1568.
2. Rogers LA (1928) The inhibiting effect of *Streptococcus lactis* on *Lactobacillus bulgaricus*. *J Bacteriol* 16(5):321-325.
3. Blin K, *et al.* (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45(W1):W36-W41.
4. van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, & Kuipers OP (2013) BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res* 41(W1):W448-W453.
5. Mohr KI, *et al.* (2015) Pinensins: the first antifungal lantibiotics. *Angew Chem Int Edit* 54(38):11254-11258.
6. Kodani S, *et al.* (2004) The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene *ramS* in *Streptomyces coelicolor*. *P Natl Acad Sci USA* 101(31):11448-11453.
7. Kodani S, Lodato MA, Durrant MC, Picart F, & Willey JM (2005) SapT, a lanthionine-containing peptide involved in aerial hyphae formation in the streptomycetes. *Mol Microbiol* 58(5):1368-1380.
8. Ferir G, *et al.* (2013) The lantibiotic peptide labyrinthopeptin A1 demonstrates broad anti-HIV and anti-HSV activity with potential for microbicidal applications. *Plos One* 8(5):e64010.
9. Iorio M, *et al.* (2014) A glycosylated, labionin-containing lanthipeptide with marked antinociceptive activity. *ACS Chem Biol* 9(2):398-404.
10. Meindl K, *et al.* (2010) Labyrinthopeptins: a new class of carbacyclic lantibiotics. *Angew Chem Int Edit* 49(6):1151-1154.
11. Repka LM, Chekan JR, Nair SK, & van der Donka WA (2017) Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chem Rev* 117(8):5457-5520.
12. Tang WX & van der Donk WA (2013) The sequence of the enterococcal cytolysin imparts unusual lanthionine stereochemistry. *Nat Chem Biol* 9(3):157-159.
13. Arnison PG, *et al.* (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30(1):108-160.
14. Goto Y, *et al.* (2010) Discovery of unique lanthionine synthetases reveals new mechanistic and evolutionary insights. *Plos Biol* 8(3):e1000339.
15. Ortega MA, *et al.* (2016) Structure and tRNA specificity of MibB, a lantibiotic Dehydratase from Actinobacteria involved in NAI-107 biosynthesis. *Cell Chem Biol* 23(3):370-380.
16. Ortega MA, *et al.* (2015) Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature* 517(7535):509.
17. Ongey EL & Neubauer P (2016) Lanthipeptides: chemical synthesis versus in vivo biosynthesis as tools for pharmaceutical production. *Microb Cell Fact* 15:97.
18. Iftime D, *et al.* (2015) Streptocollin, a Type IV lanthipeptide produced by Streptomyces collinus Tu 365. *Chembiochem* 16(18):2615-2623.

19. Zhang Q, Doroghazi JR, Zhao XL, Walker MC, & van der Donk WA (2015) Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in Actinobacteria. *Appl Environ Microb* 81(13):4339-4350.
20. Goto Y, Okesli A, & van der Donk WA (2011) Mechanistic studies of Ser/Thr dehydration catalyzed by a member of the LanL lanthionine synthetase family. *Biochemistry-Us* 50(5):891-898.
21. Hegemann JD & van der Donk WA (2018) Investigation of substrate recognition and biosynthesis in Class IV lanthipeptide systems. *J Am Chem Soc* 140(17):5743-5754.
22. Doroghazi JR, *et al.* (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* 10(11):963-968.
23. Ren HQ, Hu PF, & Zhao HM (2017) A plug-and-play pathway refactoring workflow for natural product research in *Escherichia coli* and *Saccharomyces cerevisiae. Biotechnol Bioeng* 114(8):1847-1854.

Chapter 5: Characterization of the Heat Stable Antifungal Factor (HSAF) Gene Cluster Reveals a Parallel Biosynthetic Mechanism

5.1 Introduction

Polyketides (PK) and nonribosomal peptides (NRP), as well as their hybrids, constitute one of the largest family of natural products with interesting biological activities (1). Thousands of PKs and NRPs are known to date and many of them are used as drugs for years or currently under clinical trials. The widely used PK antibiotics include erythromycin and tetracycline while the NRP antibiotics include penicillin and cephalosporin families (2-5). There are also many medicinally relevant hybrids such as the immunosuppressant drugs FK506 and rapamycin (6, 7). Although both PKs and NRPs are structurally and functionally diverse, they all arise from combinatorial utilization and elongation of simple building blocks. Malonyl-CoA and methylmalonyl-CoA are two major monomers used for PK elongation and the canonical and noncanonical amino acids are used for NRP biosynthesis. While the elongation is catalyzed by the polyketide synthase (PKS), the nonribosomal peptide synthase (NRPS) or the PKS-NRPS hybrid, tailing enzymes that are usually observed in the pathway can introduce more structural features to the final products, and some of these features are critical to the bioactivity. Based on the understanding of their biosynthesis, many efforts have been spent on developing artificial PKs, NRPs, and the hybrids by combinatorial approaches (8). Therefore, characterization of the biosynthetic mechanisms of unknown pathways for PKs, NRPs and their hybrids is of great interest.

Polycyclic tetramate macrolactams are a group of PK-NRP hybrids which contain one tetramic acid and a polycyclic system (usually 2-3 rings) which exhibit diverse biological activities such as antifungal, antibacterial and antioxidant properties (9-12). The potential PTM compounds were

first reported in 1985, when two isomers named catacandins A and B were isolated from *Lysobacter gummosus* ATCC 39742. However, the structure of PTM remain to be elusive until the 1990s, when two PTMs, named discodermide and clindramide, were isolated from sponges with their structures fully elucidated (13, 14). Recently, more PTMs were discovered from various origins, such as maltophilin from *Stenotrophomonas maltophilia* (11), heat-stable antifungal factor (HSAF) from *Lysobacter enzymogenes* (12, 15) as well as ikarugamycin (16) , dihydromaltophilin (10), and frontalamides from Streptomyces sp (9).

The heat-stable antifungal factor (HSAF) was one of the PTMs discovered from *L. enzymogenes* strain C3, which is a biological control agent originally isolated from grass foliage (17). This strain was demonstrated to reduce diseases caused by multiple fungal pathogens and the search for antifungal factors then lead to the isolation of HSAF. HSAF was demonstrated to be active against a broad range of fungal pathogens and its mode of action was studied using *Aspergillus nidulans* as a test organism (18). The polarized growth of fungus was disrupted by HSAF and further analysis of genetic mutants of *Aspergillus nidulans* suggested that HSAF targets the biosynthesis of sphingolipids. Another study also revealed that HSAF can induce the accumulation of intermediates in sphingolipid biosynthesis, which can lead to thickening of the cell wall and thus increased chitin deposition and irreversibly blocking elongation of the hyphal tips (19). Interestingly, only a distinct group of sphingolipids which is observed in filamentous fungi but absent from mammals and plants appears to be the target of HSAF, making HSAF a potential antifungal drug candidate.

Although HSAF as well as other PTMs exhibit interesting bioactivities, their biosynthetic mechanisms remain to be elusive. The proposed mechanism for PTM biosynthesis starts from the backbone formation catalyzed by a PKS-NRPS hybrid, which incorporates 12 malonyl-CoA monomers and one tetramic acid. Later, the carbon chain was modified by a series of oxidoreductases which introduce the polycyclic system as well as other groups such as hydroxyl and epoxide groups. In 2013, Zhao and coworkers successfully activated a silent PTM gene cluster from *Streptomyces griseus* by pathway refactoring. Through constructing the pathway with different gene knockouts, they proposed a biosynthetic route with multiple branches (20). While the PKS-NRPS hybrid together with two oxidoreductases in the pathway catalyze the formation of a key intermediate compound **c**, the other three enzymes in this pathway, including one oxidoreductase, one cytochrome P450 and one hydroxylase, further turn this key intermediate into three different products. The parallel biosynthetic mechanism is intriguing as in most of the cases one gene cluster generates only one final product. Therefore, it would be interesting to further investigate if such a parallel biosynthetic mechanism is widely distributed among other PTM pathways.

In this work, the biosynthetic mechanism of HSAF was investigated. Four constructs that are supposed to make different intermediates in the HSAF biosynthesis were built by DNA assembler (21). These constructs were then transformed into *Streptomyces lividans* 66 for expression and three intermediates were isolated with their structure characterized by tandem mass spectrometry (MS/MS) or nucleic magnetic resonance (NMR). The structural information clearly supports the assumption that the HSAF pathway also has a parallel biosynthetic mechanism. In addition, the distribution of PTM pathways among different microorganisms were also analyzed. Two short

PTM pathway representatives were also refactored for heterologous expression in different Streptomyces strains, which led to poor growth of the host and implied the generation of compounds that has antimicrobial activity.

5.2 Results and Discussion

5.2.1 Refactoring the HSAF Pathway to Produce dOH-HSAF in *Streptomyces lividans*

The HSAF pathway consists of six genes, including one hydroxylase gene, one PKS-NRPS hybrid gene and four oxidoreductase genes. While the activity of the hydroxylase gene has already been characterized *in vitro*, only the other five genes were refactored (22). In order to confirm if all of the five genes can be successfully expressed in *Streptomyces lividans*, they were refactored by DNA assembler as described previously which generated Construct D (Figure 5.1). Construct D was then transformed into *S. lividans* 66 through conjugation and expressed for seven days on ISP2 agar plate under 30 °C. The product was then extracted by ethyl acetate and concentrated for high performance liquid chromatography-mass spectrometry (HPLC-MS) analysis. A peak under UV 280nm was observed around 28 min with the corresponding m/z of 497.3014, which is identical to the molecular weight of the reported deOH-HSAF (calculated m/z: 497.2971, an intermediate in HSAF biosynthesis obtained through knockout the hydroxylase in the native producer (Figure 5.2).

5.2.2 Design of Partially Refactored HSAF Pathways

As all of the five genes under investigation were shown to be active in the heterologous host, we proceeded to design constructs with other gene combinations to further investigate the biosynthetic mechanism. Sequence analysis of the four oxidoreductases (OX1-OX4) in the PTM pathway from

S. griseus showed that both OX1 and OX2 are homologous to SGR813, OX3 is homologous to SGR812 and OX4 is homologous to SGR811, while the SGR810 homolog is not observed in the HSAF pathway. Based on the proposed biosynthetic mechanism for the PTM pathway in *S. griseus*, another three constructs were designed as follows: Construct A contains the PKS-NRPS hybrid with OX2 and OX3 which is supposed to generate the key intermediate; OX1 and OX4 were added to Construct A respectively which generate Construct B and Construct C (Figure 5.1). If the HSAF pathway follows the parallel mechanism as assumed, Construct B and Construct C should produce different compounds.

5.2.3 Characterization of the Compounds Produced by the Partial HSAF Pathways

Construct A/B/C were built by DNA assembler and transformed into *S. lividans* 66 for expression as described previously. The products were also extracted and analyzed by the same approach used for dOH-HSAF. Initial HPLC-MS analysis indicated that the 3-gene cluster of the HSAF pathway (Construct A) produced an intermediate with a corresponding m/z of 477.2815, which is nearly identical to a proposed intermediate with a single ring. The chemical structure was subsequently confirmed by 1D and 2D NMR (COSY, HSQC and HMBC) (Figure 5.3 and Figure 5.4). The 4-gene HSAF cluster (Construct B) produced an intermediate with the same retention time with the deOH-HSAF produced by the five-gene HSAF cluster (Construct D). The corresponding m/z of 497.3014 further confirmed that the product from this four-gene cluster is deOH-HSAF. Another 4-gene HSAF cluster (HSAF-C) revealed the production of an intermediate with a relatively longer retention time compared to deOH-HSAF and a corresponding m/z of 495.2853 (Figure 5.2). Compared with the MS/MS data of previously characterized compounds, this intermediate was identified as compound c, which is the key intermediate in the PTM pathway from *S. griseus*.

5.2.4 Analysis of the Distribution of the PTM Pathways and Characterization of the PTM Pathways from *Saccharophagus degradans* and *Salinospora arenicola*

In order to further investigate if the parallel biosynthetic mechanism exists in PTM biosynthetic gene clusters, over 800 sequenced actinobacterial genomes were analyzed for the existence of PTM gene cluster homologs (Figure 5.5). Most of these gene clusters reveal similar combination of biosynthetic genes as the HSAF pathway and the one previously discovered from *S. griseus*. Besides the five-gene clusters, three-gene clusters and six-gene clusters were also observed in some strains, which may produce PTMs with different structural features from the known ones.

As there is no report for any PTMs produced from these short gene clusters, two of them were picked for characterization by using the same approach for the HSAF cluster (Figure 5.5). The cluster from *Saccharophagus degradans* contains only three genes with only one oxidoreductase which is the homolog to SGR812. The cluster from *Salinospora arenicola* contains four genes which resemble the gene cluster for ikarugamycin biosynthesis. However, after refactoring and transformation into multiple Streptomyces strains for expression (*S. lividans* 66, *S. albus* J1074, *S. coelicolor* M1146, and *S. coelicolor* M1152), all hosts showed a very poor cell growth rate and thus no product was detected. However, such a phenomenon implied that the potential products from these two pathways might have antimicrobial activity.

5.2.5 Proposed Mechanism of PTM Biosynthesis

The results from the four HSAF constructs suggest a parallel biosynthetic mechanism of the HSAF pathway. The OX2 and OX3 are critical for the synthesis of the key intermediate **1**. While OX4

can further convert this intermediate into a three-ring product deOH-HSAF **2**, OX1 can convert it into a two-ring product compound **3** (Figure 5.6). The formation of the bicyclic scaffold or the tricyclic scaffold is determined by the differential expression of two critical oxidoreductases involved in these PTM biosynthetic gene clusters. Bioinformatics analysis of over 800 sequenced actinobacteria genomes revealed the wide distribution of the PTM biosynthetic pathways, which indicates the general presence of this parallel biosynthesis mechanism in actinobacteria. This mechanism might be evolutionarily advantageous when generating a new natural product biosynthetic pathway because several genes can be shared and fewer mutational steps are required to meet a new selective pressure. In fact, there are PTM biosynthetic pathways only consisting of three genes, which may be the minimal set to produce a desired natural product against environmental pressure. For example, the PTM gene cluster from *Saccharophagus degradans* only contains homologs of SGR812, SGR815 and SGR814, which supports the proposed role of SGR812 in the first ring formation and compound **1** as the key intermediate. As in *Salinospora arenicola* and the ikarugamycin producing strain, the SGR811 homologs are found downstream of the SGR812 homolog, while no SGR813 homologs are found. Based on the structure of ikarugamycin, the function of SGR811 homologs can be proposed as forming the inner two-ring structure.

In addition, there are PTM pathways with six genes, which may be able to produce two sets of compounds to deal with different selection pressure. Taken the PTM pathway from *S. griseus* as an example, SGR811 and SGR813 are the two key enzymes tuning the expression of this pathway into two branches. These two enzymes are annotated as oxidoreductases and can be found in most of the known PTM gene clusters. Based on phylogenetic analysis, these two enzymes are distinct

and this is probably why they could catalyze different reactions even though they belong to the same family of oxidoreductases. If these 6-gene containing PTM pathways are evolved from the 3-gene pathways, the evolutionary changes presented here would be quantitative, not qualitative. The quantitative changes are important in evolutionary diversification. In the PTM evolution, rather than incorporating a promiscuous enzyme for new compounds, one additional gene is inserted into an existing pathway to generate diversity. To our knowledge, this is by far the first example of multiple natural products generated from the same gene cluster by the tuning of two key genes. More importantly, this phenomenon is prevailing in nature because the PTM biosynthetic machinery is well preserved.

5.3 Conclusion

In this work, the biosynthesis of HSAF was characterized by pathway refactoring and heterologous expression in *S. lividans*. Structural characterization of the intermediates produced by different partial HSAF pathways reveal a parallel biosynthetic route that is similar to that in the PTM pathway discovered from *S. griseus*. Bioinformatic analysis indicates that this mechanism may be a general mechanism in PTM biosynthesis.

5.4 Materials and Methods

5.4.1 Strains and Materials

Saccharomyces cerevisiae HZ848 (*MATα, ade2-1, Δura3, his3-11, 15, trp1-1, leu2-3, 112 and can1-100*) was used as the host for DNA assembly. *S. lividans* 66 was obtained from the Agricultural Research Service Culture Collection (Peoria, IL). *Lysobacter enzymogenes* strain C3 was a gift from Dr. Liangcheng Du (University of Nebraska-Lincoln, NE). Plasmids pAE4, *E. coli*

strain BW25141 [*lacIq , rrnBT14, ΔlacZWJ16, ΔphoBR580, hsdR514, ΔaraBADAH33, ΔrhaBADLD78, galU95, endABT333, uidA(ΔMluI)∴ pir+ recA1*, derived from *E. coli* K-12 strain BD792] and *E. coli* strain WM6026 [*lacIq, rrnB3, ΔlacZ4787, hsdR514, ΔaraBAD567, Δ rhaBAD568, rph-1, attl::pAE12 (Δ oriR6K-cat::Frt5), Δ endA::Frt, uidA(Δ MluI)::pir, attHK::pJK1006 Δ (oriR6K-cat::Frt5; trfA::Frt)*] were provided by Dr. William Metcalf (University of Illinois, Urbana, IL). The plasmid pRS416 was purchased from New England Biolabs (Ipswich, MA). Nalidixic acid and Isopropyl ß-D-1-thiogalactopyranoside (IPTG) were obtained from Sigma-Aldrich (St. Louis, MO). ISP2, agar, beef extract, yeast extract, malt extract, and other reagents required for cell culture were obtained from Difco (Franklin Lakes, NJ). All restriction endonucleases and Q5 DNA polymerase were purchased from New England Biolabs. The QIAprep Spin Plasmid Mini-prep Kit, QIAquick PCR Purification Kit, QIAquick Gel Extraction Kit were purchased from Qiagen (Valencia, CA). Wizard Genomic DNA isolation kit was from Promega (Madison, WI). Zymoprep II Yeast Plasmid Miniprep kit was purchased from Zymo (Zymo Research, CA). All primers were synthesized by Integrated DNA Technologies (Coralville, IA). Yeast YPAD medium containing 1% yeast extract, 2% peptone and 2% dextrose supplemented with 0.01% adenine hemisulphate was used to grow *S. cerevisiae* HZ848. Synthetic complete drop-out medium lacking uracil (SC-Ura) was used to select transformants containing the assembled biochemical gene clusters of interest. MYG liquid medium containing 4% yeast extract, 10% malt extract, 4% glucose and YMS solid agar medium containing 4% yeast extract, 10% malt extract, 4% soluble starch, and 20% Bacto agar were used to grow Streptomyces strains.

5.4.2 Gene Cluster Reconstruction and Yeast Transformation

Gene cluster fragments were amplified from the genomic DNA of *L. enzymogenes* strain C3, respectively. The *S. cerevisiae* helper fragment was amplified from the plasmid pRS416, whereas the *E. coli* helper fragment and the *S. lividans* helper fragment were amplified from pAE4 (23). The PCR products were individually gel-purified from 1% agarose. Promoters from other hosts were obtained through previous studies (24, 25). Genes with corresponding promoters were ligated together by overlap extension PCR. To ensure high efficiency of yeast homologous recombination, the "promoter upstream-gene-promoter downstream" cassette was built to generate a ~200 bp overlap region with adjacent fragments. The cassettes were mixed and concentrated, then electroporated into *S. cerevisiae* following the DNA assembler protocol (21).

5.4.3 Heterologous Expression in *S. lividans*

The verified clones were transformed to *E. coli* WM6026 and selected on LB agar plates supplemented with 19 μg/mL 2,6-diaminopimelic acid and 50 μg/mL apramycin. These transformants were then used as the donors for conjugative transfer of the assembled plasmids to *S. lividans* 66 following the protocol described elsewhere (26). *S. lividans* exconjugants were picked and re-streaked on ISP2 plates supplemented with 50 μg/mL apramycin and grown for 2 days. Then a single colony was inoculated into 15 mL of MYG liquid medium supplemented with 50 μg/mL apramycin in a 125 mL shake-flask (3 mm glass beads added to improve liquid mixing and aeration). Aliquots of this seed culture with a volume of 250 μL were spread for confluence over Petri plates containing approximately 25-30 mL of YMS solid agar medium (4 g/L yeast extract, 10 g/L malt extract, 4 g/L soluble starch, and 20 g/L Bacto agar) and the plates were incubated for 7 days at 30 °C.

5.4.4 HPLC-MS Analysis

To isolate potential compounds, agar and the adherent cells were stored at -80 °C for at least 24 h, after which the frozen cultures were thawed completely and the released liquids were collected. The liquids were extracted by ethyl acetate with a ratio of 1:1 twice, concentrated 1000-fold, and subjected to HPLC analysis. HPLC was performed on the Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer with a Phenomenex Luna C18 reverse-phase column (3.0 x 150 mm, 3.5 μm). HPLC parameters were as follows: solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile; gradient, 10% B for 5 min, to 100% B in 25 min, maintain at 100% B for 10 min, return to 10% B in 1 min and finally maintain at 10% B for 9 min; flow rate 0.3 mL/min; detection by UV spectroscopy at 260 nm and 400 nm. The MS system was operated using a drying temperature of 350 C, a nebulizer pressure of 35 psi, a drying gas flow of 8.5 L/min, and a capillary voltage of 4500 V.

5.4.5 Structure Elucidation of Compound 1

Compound **1** was isolated by reverse-phase HPLC using a Phenomenex Luna 10u C18 column (10 mm × 250 mm), on the Agilent 1100 series LC/MSD XCT with an isocratic solvent of 60% acetonitrile in water with 0.1% formic acid (with a flow rate of 3 mL/min and a wavelength of 260 nm). The high resolution MS was performed on a Waters Q-Tof Ultima system (Walters, Milford, MA). All NMR experiments were carried out on a Varian UNITY INOVA 600 MHz spectrometer, except the ^{13}C spectra, which were recorded on a Bruker Avance III 500i with a 5 mm $^{13}C(^{1}H)$ cryogenic probe at the National Magnetic Resonance Facility at the University of Wisconsin at Madison. The structures were determined based on 1-D and 2-D NMR data.

5.5 Figures

Figure 5.1. Plasmid maps from Construct A to Construct D.

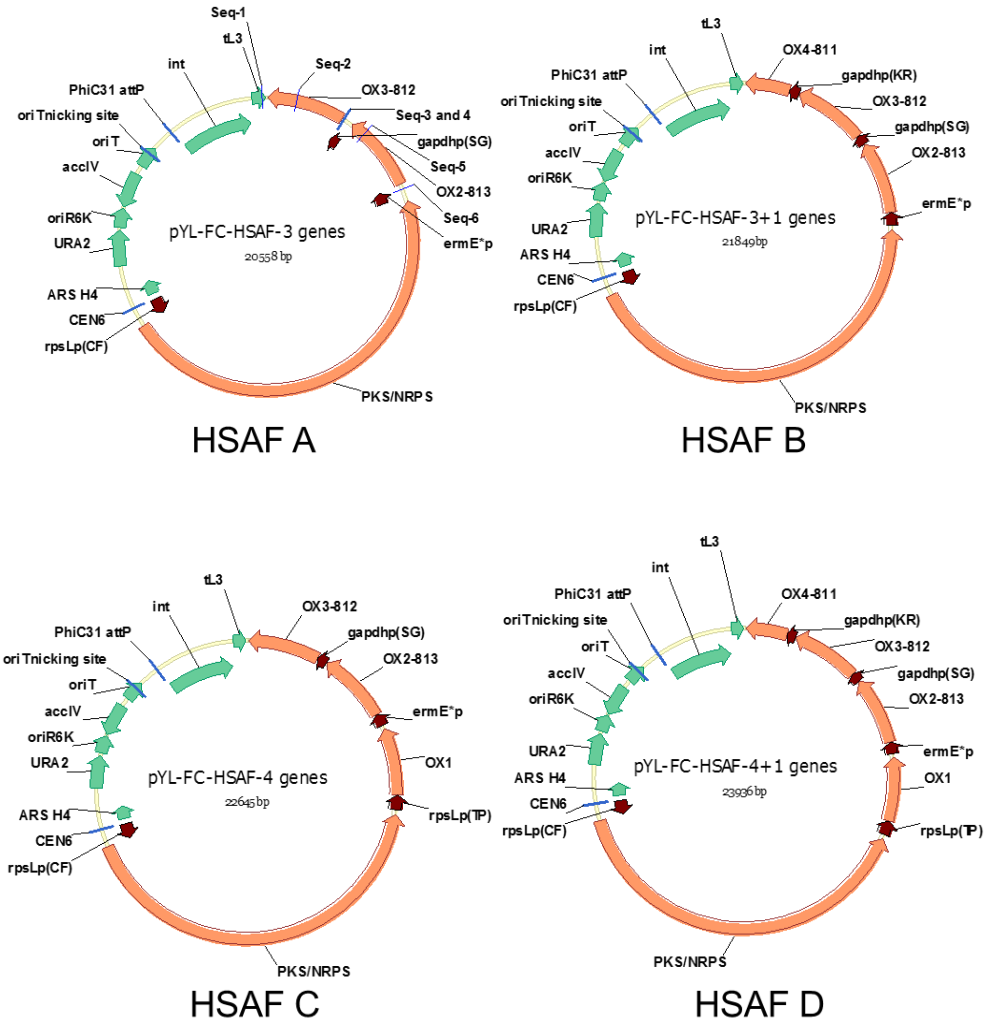


Figure 5.2. UV 280nm traces for crude extracts from *S. lividans* 66 with Construct A/B/C/D on HPLC. Asterisks indicate the peak with structural information.

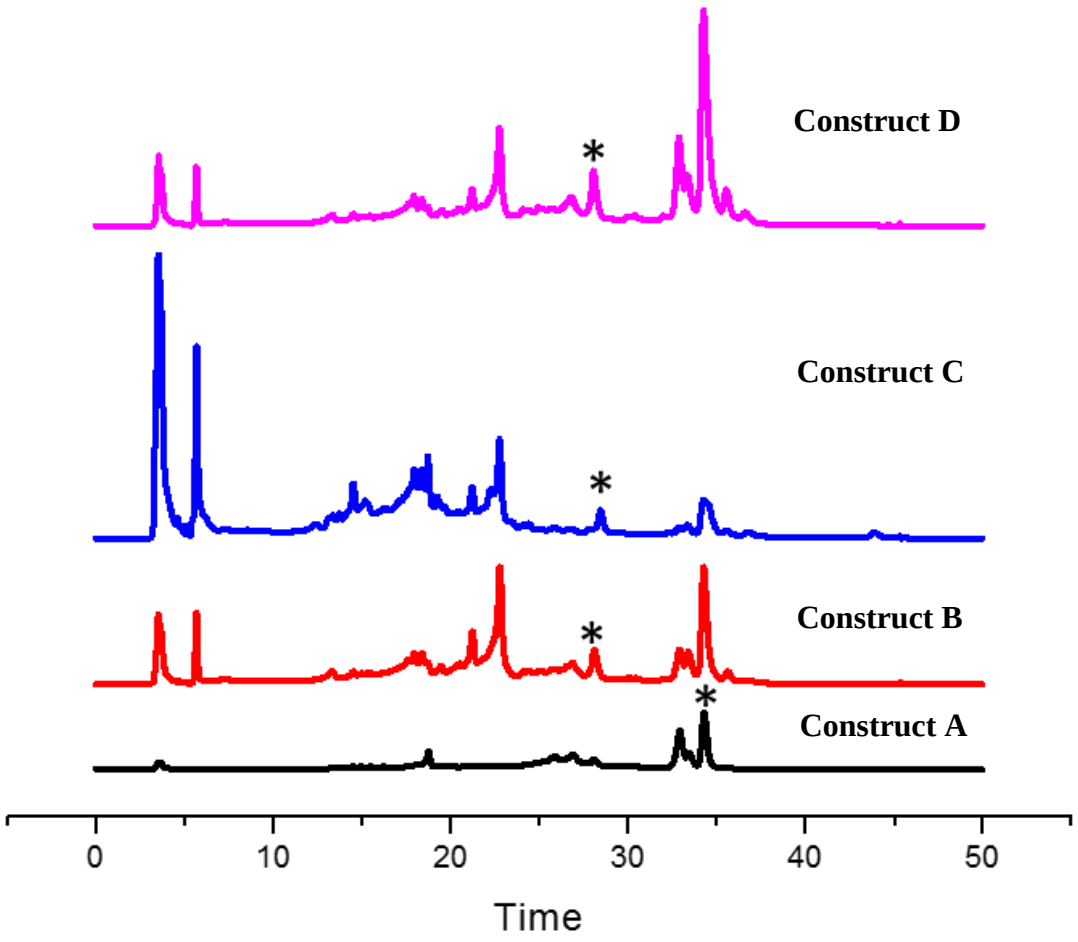
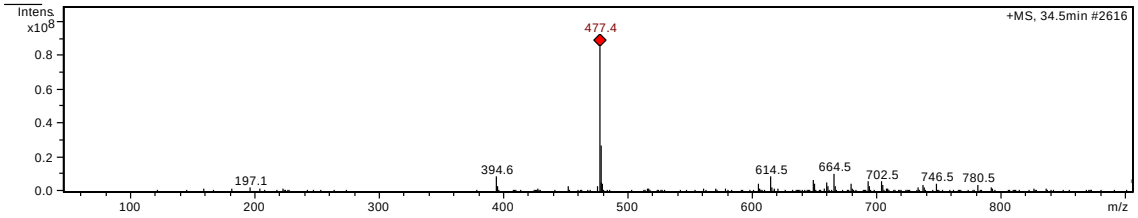


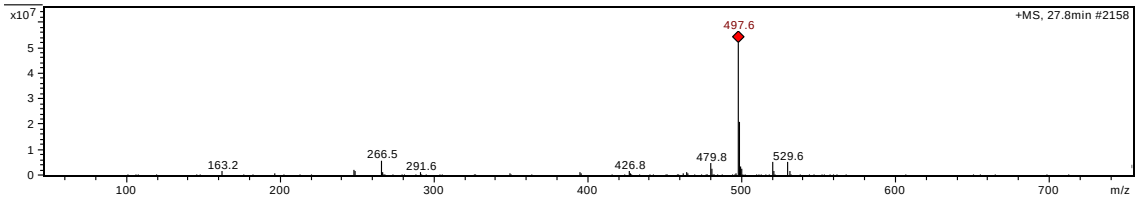
Figure 5.3. MS and MS/MS results for compound **1-3**.

MS

Compound **1**



Compound **2**



Compound **3**

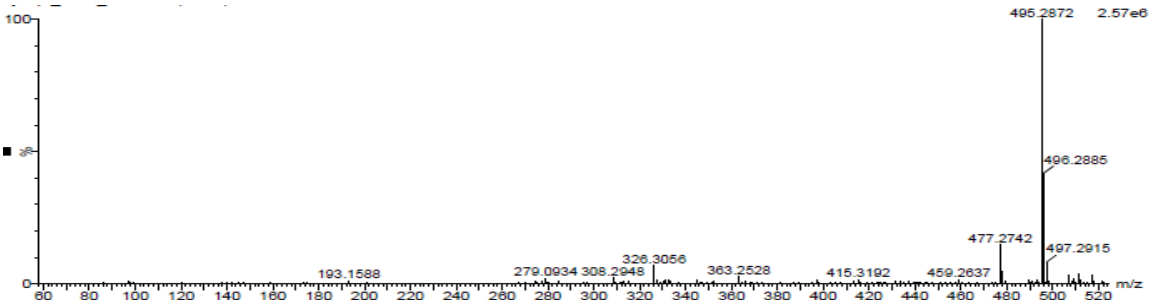


Figure 5.3. Continue

MS/MS

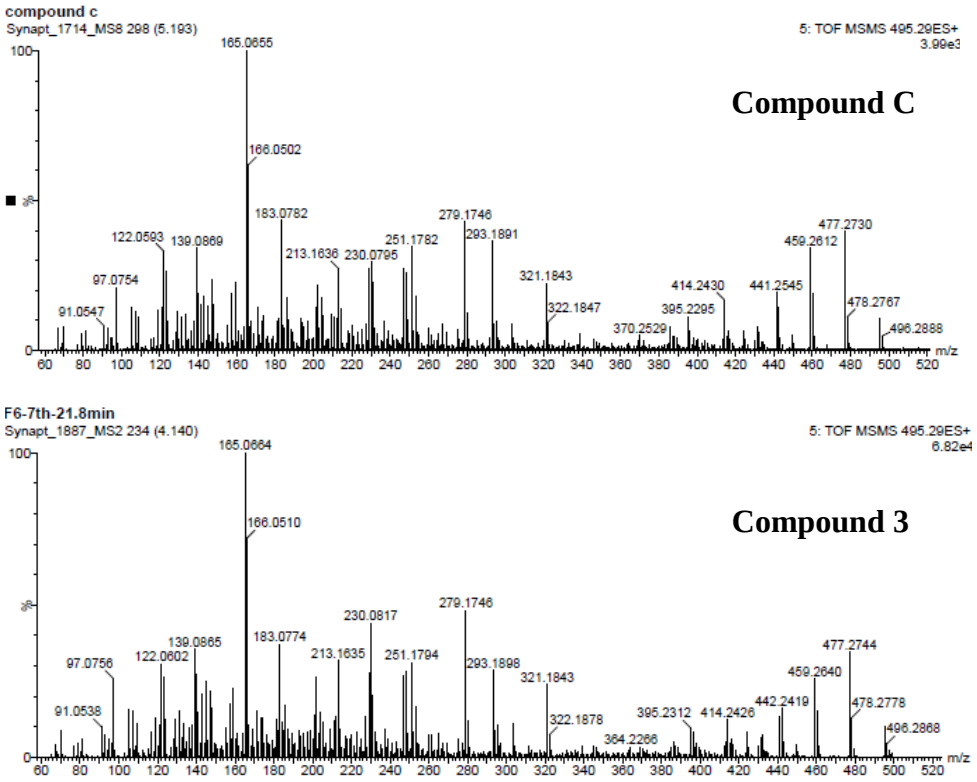


Figure 5.4. NMR results for compound **1**.

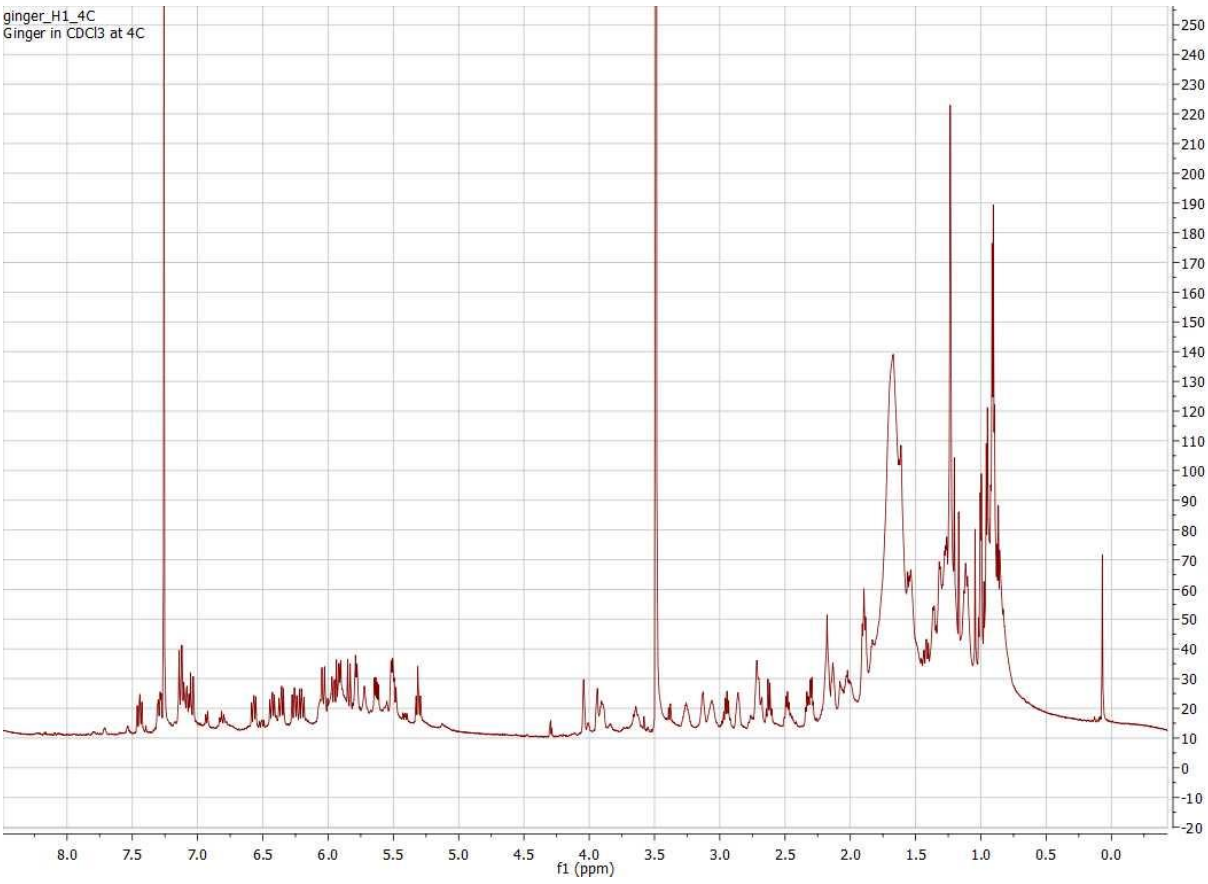


Figure 5.4. Continue

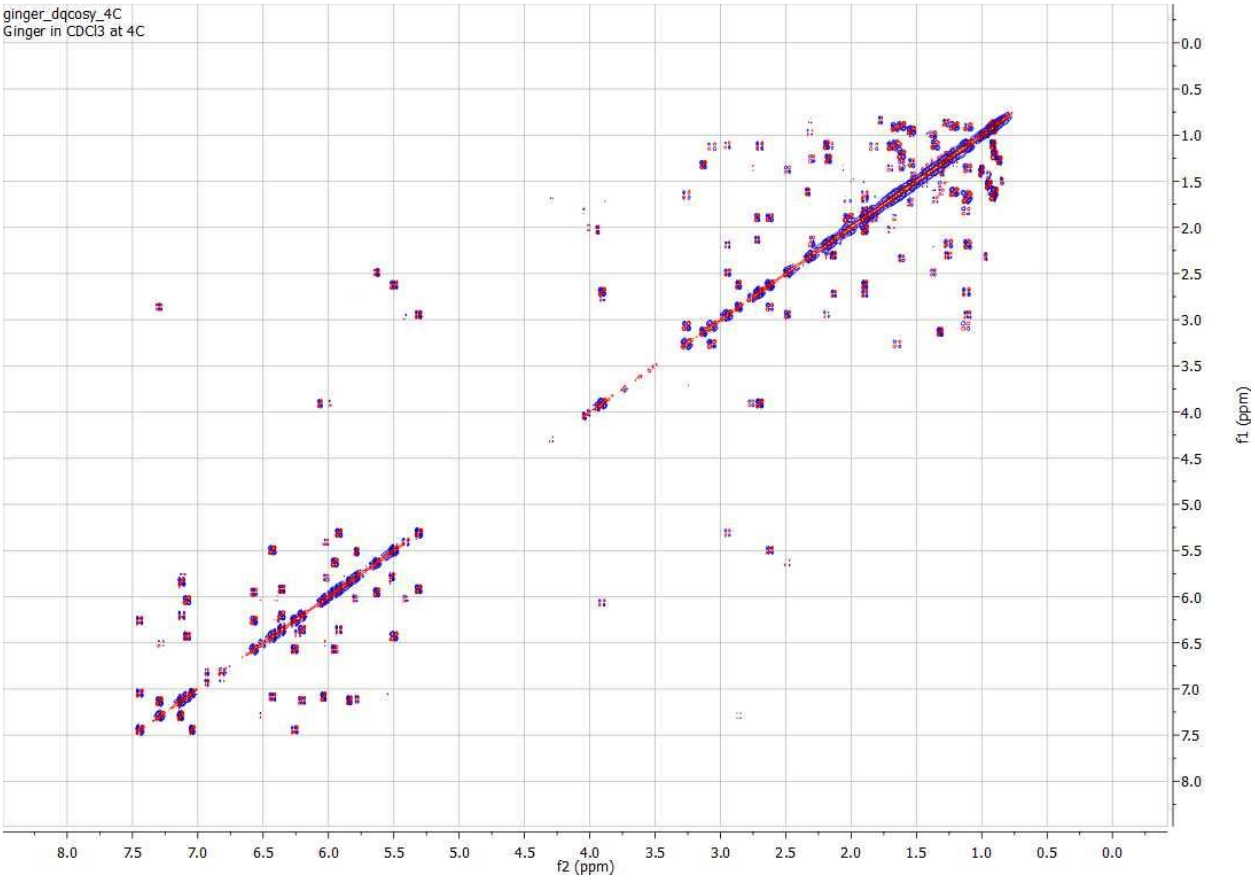


Figure 5.4. Continue

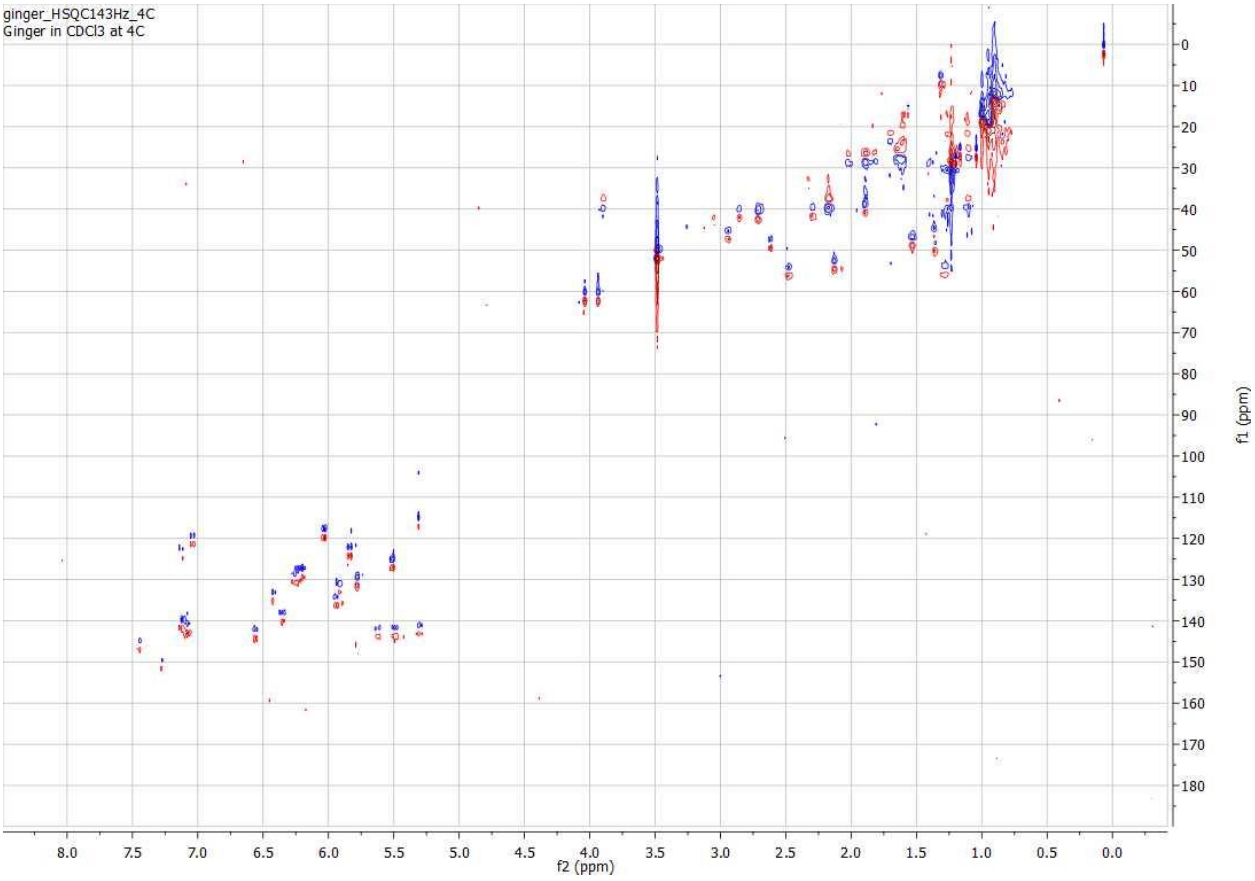


Figure 5.4. Continue

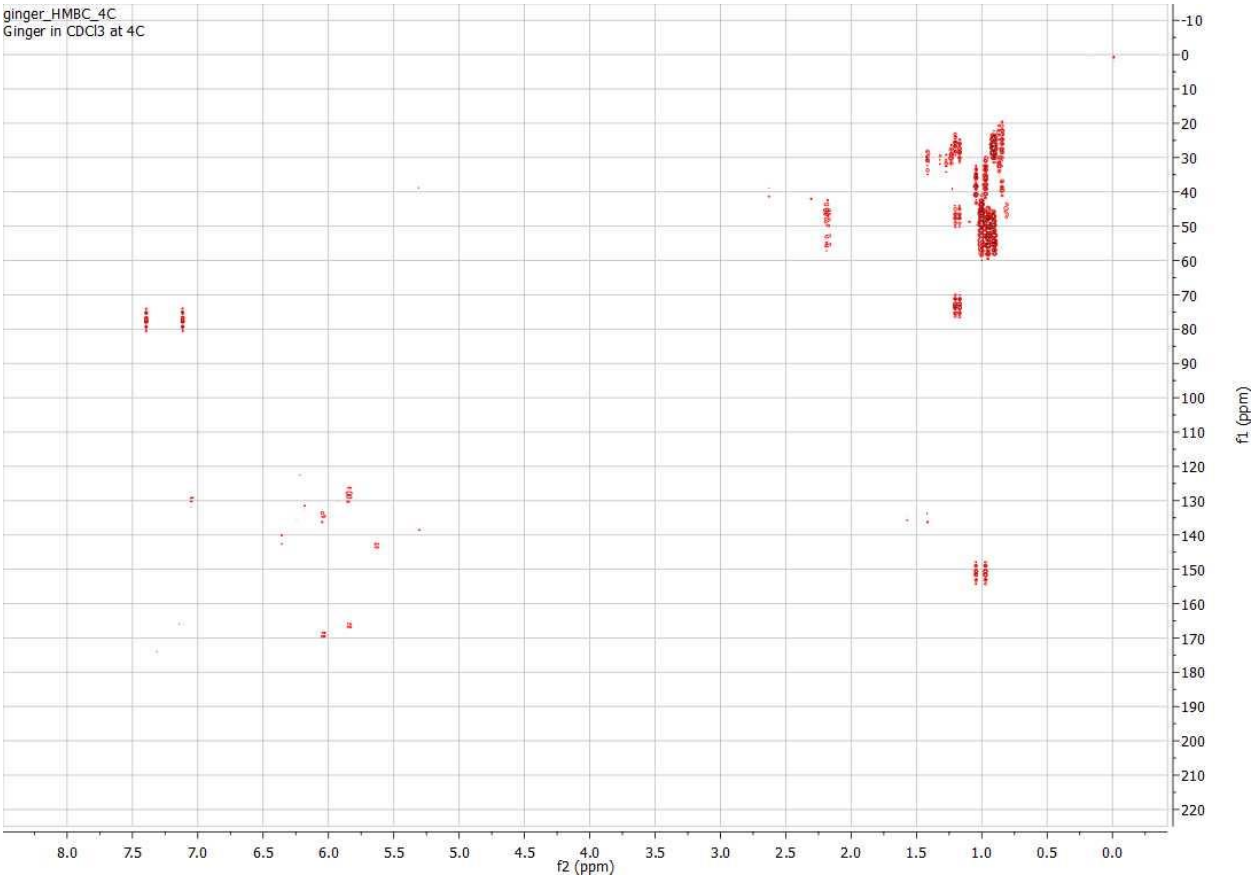


Figure 5.5. Representatives of PTM biosynthetic gene clusters. 1. SGR815 homologs; 3: SGR813 homologs; 4: SGR812 homologs; 5: SGR811 homologs; 6: SGR810 homologs.

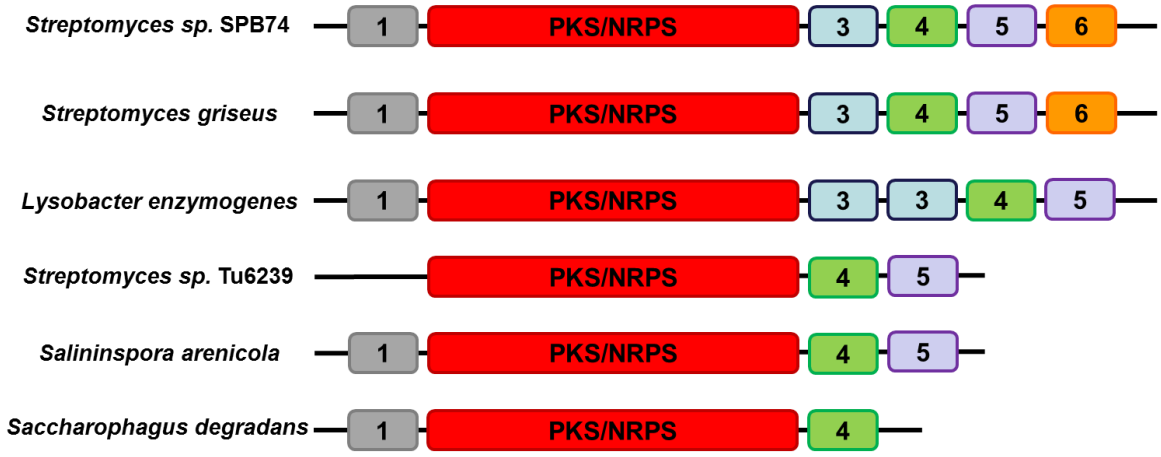
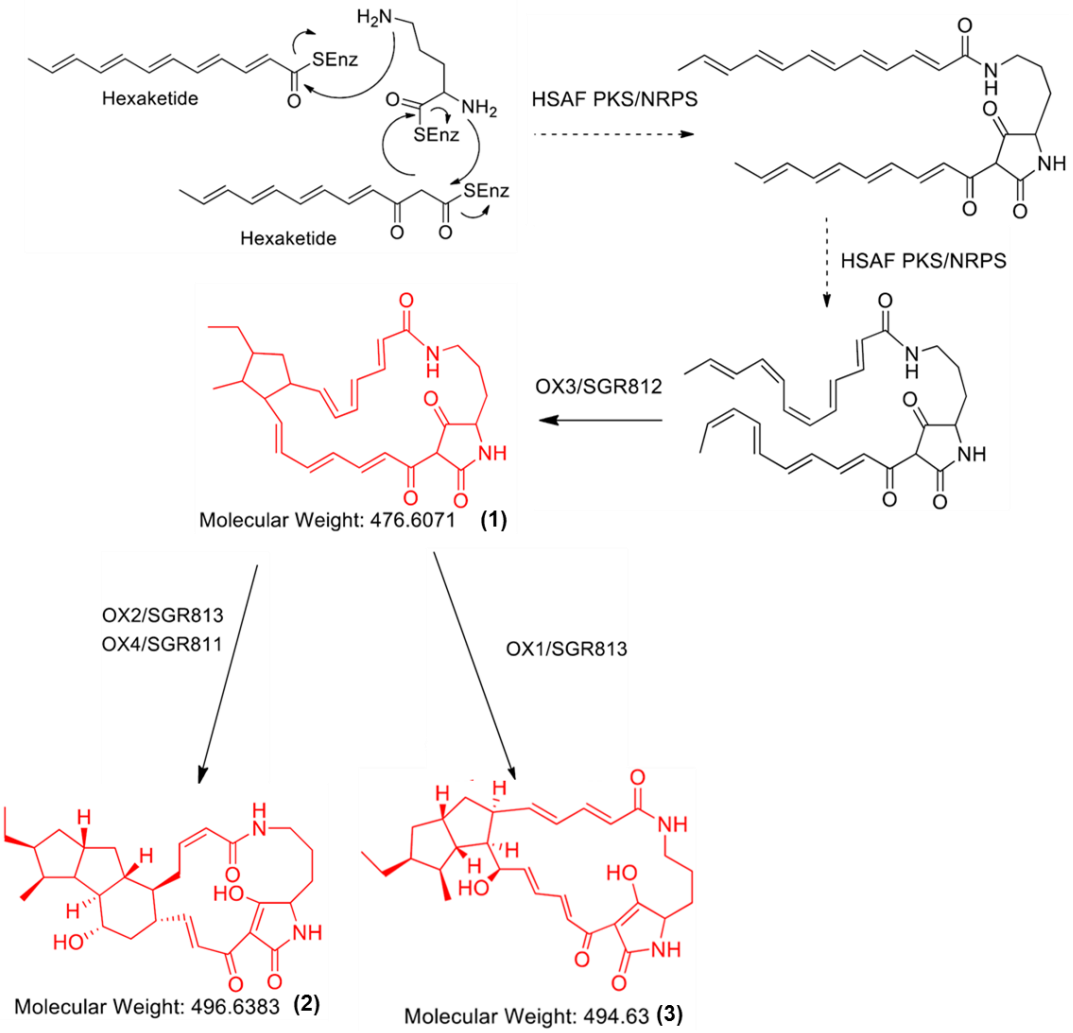


Figure 5.6. Proposed biosynthetic mechanism of HSAF.



5.6 References

1. Fischbach MA & Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem Rev* 106(8):3468-3496.
2. Donadio S, Staver MJ, Mcalpine JB, Swanson SJ, & Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science* 252(5006):675-679.
3. Kim ES, Bibb MJ, Butler MJ, Hopwood DA, & Sherman DH (1994) Sequences of the oxytetracycline polyketide synthase-encoding otc genes from Streptomyces rimosus. *Gene* 141(1):141-142.
4. Aharonowitz Y, Cohen G, & Martin JF (1992) Penicillin and cephalosporin biosynthetic genes - structure, organization, regulation, and evolution. *Annu Rev Microbiol* 46:461-495.
5. Martin JF (1998) New aspects of genes and enzymes for beta-lactam antibiotic biosynthesis. *Appl Microbiol Biot* 50(1):1-15.
6. Motamedi H & Shafiee A (1998) The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506. *Eur J Biochem* 256(3):528-534.
7. Schwecke T, *et al.* (1995) The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *P Natl Acad Sci USA* 92(17):7839-7843.
8. Cai WL & Zhang WJ (2018) Engineering modular polyketide synthases for production of biofuels and industrial chemicals. *Curr Opin Biotech* 50:32-38.
9. Blodgett JAV, *et al.* (2010) Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *P Natl Acad Sci USA* 107(26):11692-11697.
10. Graupner PR, *et al.* (1997) Dihydromaltophilin: a novel fungicidal tetramic acid containing metabolite from Streptomyces sp. *J Antibiot* 50(12):1014-1019.
11. Jakobi M, *et al.* (1996) Maltophilin: a new antifungal compound produced by *Stenotrophomonas maltophilia* R3089. *J Antibiot* 49(11):1101-1104.
12. Yu FG, *et al.* (2007) Structure and biosynthesis of heat-stable antifungal factor (HSAF), a broad-spectrum antimycotic with a novel mode of action. *Antimicrob Agents Chemother* 51(1):64-72.
13. Meyers E, *et al.* (1985) Catacandins, novel anticandidal antibiotics of bacterial origin. *J Antibiot* 38(12):1642-1648.
14. Gunasekera SP, Gunasekera M, & Mccarthy P (1991) Discodermide - a new bioactive macrocyclic lactam from the marine sponge *Discodermia dissoluta*. *J Org Chem* 56(16):4830-4833.
15. Lou LL, *et al.* (2011) Biosynthesis of HSAF, a tetramic acid-containing macrolactam from *Lysobacter enzymogenes*. *J Am Chem Soc* 133(4):643-645.
16. Greunke C, Glockle A, Antosch J, & Gulder TAM (2017) Biocatalytic total synthesis of ikarugamycin. *Angew Chem Int Edit* 56(15):4351-4355.
17. Giesler LJ & Yuen GY (1998) Evaluation of *Stenotrophomonas maltophilia* strain C3 for biocontrol of brown patch disease. *Crop Prot* 17(6):509-513.
18. Li SJ, Du LC, Yuen G, & Harris SD (2006) Distinct ceramide synthases regulate polarized growth in the filamentous fungus *Aspergillus nidulans*. *Mol Biol Cell* 17(3):1218-1227.

19. Li SJ, Calvo AM, Yuen GY, Du LC, & Harris SD (2009) Induction of cell wall thickening by the antifungal compound dihydromaltophilin disrupts fungal growth and is mediated by sphingolipid biosynthesis. *J Eukaryot Microbiol* 56(2):182-187.
20. Luo YZ, *et al.* (2013) Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nat Commun* 4:2894.
21. Shao ZY, Zhao H, & Zhao HM (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* 37(2):e16.
22. Li YY, Huffman J, Li Y, Du LC, & Shen YM (2012) 3-Hydroxylation of the polycyclic tetramate macrolactam in the biosynthesis of antifungal HSAF from *Lysobacter enzymogenes* C3. *Medchemcomm* 3(8):982-986.
23. Eliot AC, *et al.* (2008) Cloning, expression, and biochemical characterization of streptomyces rubellomurinus genes required for biosynthesis of antimalarial compound FR900098. *Chem Biol* 15(8):765-770.
24. Shao ZY, *et al.* (2013) Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth Biol* 2(11):662-669.
25. Luo YZ, Zhang L, Barton KW, & Zhao HM (2015) Systematic identification of a panel of strong constitutive promoters from *Streptomyces albus*. *ACS Synth Biol* 4(9):1001-1010.
26. Woodyer RD, *et al.* (2006) Heterologous production of fosfomycin and identification of the minimal biosynthetic gene cluster. *Chem Biol* 13(11):1171-1182.

Chapter 6: Development of CRISPR-based Genetic Tools for Natural Product Discovery and Overproduction in Streptomyces Species

6.1 Introduction

Streptomycetes are filamentous, Gram-positive bacteria with high GC content genomes that are predominantly found in soil and water environments. The most notable property of this genus is their ability to produce pharmaceutically relevant secondary metabolites including antimicrobial, antifungal, antiviral, antihypertensive, immune suppressant and antitumor activities (1). In 2002, when the genome of a streptomyces species, *Streptomyces coelicolor* A3(2), was fully sequenced for the first time, the authors identified more than 20 biosynthetic gene clusters (BGCs) encoding secondary metabolites (2). Although some of them were already known, others had not yet been identified. Since then, an increasing number of streptomyces species were sequenced and each encodes dozens of BGCs. Besides being a prolific resource for discovery of new compounds, streptomycetes are also ideal chassis for overproducing natural products. Adequate substrates along with multiple post-translational modification systems are available in streptomycetes, which are necessary for natural product biosynthesis. To date, metabolic engineering approaches have been successfully used for optimizing the productivity of some pharmaceutically important natural products in a few Streptomyces species (3). However, the existing genetic tools for engineering streptomycetes are generally time-consuming and inefficient.

Clustered regularly interspaced short palindromic repeats (CRIPSR)-Cas9 system has been developed as a powerful genetic engineering tool for both eukaryotes and prokaryotes (4, 5). The most widely used Type II CRISPR-Cas9 system from *Streptococcus pyogenes* consists of the Cas9 nuclease and a chimeric single guide RNA (sgRNA). The 20 nt protospacer sequence in sgRNA

can help Cas9 specifically recognize the designated DNA sequence and perform double strand cleavage. Because of its excellent specificity in target recognition, the CRISPR-Cas9 system was further developed into CRISPR interference (CRISPRi) system (6). The CRISPRi system consists of a nuclease deficient Cas9 (dCas9) which can recognize the target sequence but cannot perform cleavage. Through targeting either the transcription initiation region or the coding region, CRISPRi can specifically and reversibly repress the expression of its target gene. Furthermore, if the dCas9 is fused with transcriptional activator and guided to the suitable site, it is able to recruit RNA polymerase (RNAP) and therefore improve the expression level of target gene, which is referred as CRISPRa (7-10). The application of CRISPR-Cas9 in several streptomycetes and CRISPRi in *S. coelicolor* were first demonstrated in 2015 by several research groups (11-14). However, there is still no published data showing the activity of CRISPRa in streptomycetes. In this chapter, I demonstrated the activity of CRISPRi system in both *Streptomyces lividans* 66 and *Streptomyces albus* J1074, which are two most widely used model Streptomyces strains as *S. coelicolor*. By fusing the dCas9 with the RNAP-ω factor, CRISPRa was also proved to be active in *S. lividans* 66 through activation of the undecylprodigiosin (RED) gene cluster. To achieve titratable repression, an inducible promoter was used for sgRNA expression. However, further optimization is needed to broaden the tunable range for expression.

6.2 Results and Discussion

6.2.1 Characterization of a CRISPRi System in *S. lividans* 66 and *S. albus* J1074

In order to characterize the activity of dCas9 in *S. lividans* 66 and *S. albus* J1074, XylE, a widely used sterptomycetes reporter that can convert the colorless substrate catechol to an intensely yellow hydroxymuconic semialdehyde, was chosen as the target gene. The *xylE* expression

cassette along with the sgRNA under the control of *gadph* promoter and *oop* terminator were cloned into the pSET-dCas9 plasmid obtained from AddGene (15). Two sgRNAs targeting both the template (xylE-T) and nontemplate strand (xylE-NT) of the coding region of *xylE* were tested. In both strains, only the sgRNA targeting the nontemplate strand exhibited significant repression (Figure 6.1). Such an observation is consistent with the results published by Tong and coworkers, in which the production of actinorhodin in *S. coelicolor* was significantly repressed only when the sgRNAs targeted the nontemplate strand in the coding region of its key biosynthetic gene *actIORF1* (13). Since the CRISPRi system showed higher level of repression in *S. lividans* 66, the following experiments were performed in *S. lividans* 66 instead of *S. albus* 1074.

6.2.2 Development and Characterization of a CRISPRa System in *S. lividans*

To further develop the CRISPRa system, the *rpoZ* gene encoding the RNAP-ω factor was identified from *S. lividans* 66 and fused to the dCas9 protein with a double alanine linker. The *rpoZ* was fused to the C-terminal instead of N-terminal of dCas9 since a similar strategy was previously demonstrated in *Escherichia coli* and the C-terminal fusion showed better performance (7). Multiple sgRNAs targeting the promoter region of *redD* gene, activator of the red pigment undecylprodigiosin biosynthesis in *S. lividans* 66 which is silent under certain culturing conditions, were designed and coexpressed with the dCas9/RNAP-ω fusion protein (Figure 6.2A). Among eight sgRNAs, only two were observed to increase the productivity of red pigment (Figure 6.2B). Similar to the CRISPRa system in other organisms, the distance between the targeting region and the transcription start site seems to be critical for its performance in streptomycetes.

6.2.3 Development and Characterization of a Titratable CRISPRi System in *S. lividans*

The ability to fine tune native gene expression is critical for metabolic engineering as in many cases intermediate expression levels result in maximum product titer, and such a task can be achieved by CRISPRi system through controlling the expression level of the dCas9-sgRNA complex (16). In order to investigate if such a strategy is also applicable in streptomycetes, an inducible promoter *Potr* identified from *Streptomyces rimousus* was used to control the sgRNA expression (Figure 6.3A) (17). Although *Potr* was characterized with a broad tunable range in multiple streptomycetes, its background activity was sufficient to lead significant repression which resulted in very limited tunable range for repression (Figure 6.3B). Therefore, further improvement is needed to reduce the background activity and increase the tunable range.

6.3 Conclusion

In this chapter I have shown some preliminary data for developing CRISPR-based genetic engineering tools for streptomycetes. The activity of CRISPRi was demonstrated in both *S. lividans* 66 and *S. albus* J1074. By fusing the *rpoZ* gene encoding the RNAP-ω factor to dCas9, the activation of silent genes was achieved by targeting the appropriate site in the promoter region. In addition, titratable repression of CRISPRi by using an inducible promoter to control the expression level of dCas9/sgRNA complex was also investigated, which requires further optimization.

6.4 Materials and Methods

6.4.1 Strains and Media

All chemicals were purchased from Sigma-Aldrich (St. Louis, MO) or ThermoFisher Scientific (Waltham, MA). *E. coli* NEB® 10-beta (New England Biolabs, Ipswich, MA) was used for DNA manipulation. *E. coli* strain ET12567/pUZ800235 was a gift from Professor Eriko Takano (The University of Manchester, Manchester, United Kingdom) and used for plasmid conjugation. *S. lividans* 66 was obtained from the Agricultural Research Service Culture Collection (Peoria, IL). *S. albus* J1074 was a gift from Professor Wenjun Zhang (University of California, Berkeley, CA). ISP2, agar, beef extract, yeast extract, malt extract, and other reagents required for cell culture were obtained from Difco (Franklin Lakes, NJ). All restriction endonucleases and Q5 DNA polymerase were purchased from New England Biolabs. The QIAprep Spin Plasmid Mini-prep Kit, QIAquick PCR Purification Kit, QIAquick Gel Extraction Kit were purchased from Qiagen (Valencia, CA). Wizard Genomic DNA isolation kit was from Promega (Madison, WI). All primers were synthesized by Integrated DNA Technologies (Coralville, IA). MYG liquid medium containing 4% yeast extract, 10% malt extract, 4% glucose, MYG solid agar medium containing additional 2% Bacto agar and MS solid agar medium containing 2% soy bean flour, 2% maltose, and 2% Bacto agar were used to grow Streptomyces strains.

6.4.2 Conjugation and Expression in Streptomyces species

The CRISPR plasmids were transformed to *E. coli* ET12567/pUZ8002 and selected on LB agar plates supplemented with 25 μg/mL apramycin, 25 μg/mL kanamycin and 12.5 μg/mL chloramphenicol. These transformants were then used as the donors for conjugative transfer of the assembled plasmids to streptomycetes following the protocol described elsewhere (18).

Exconjugants were picked and re-streaked on MS plates supplemented with 50 µg/mL apramycin and grown for 3 days. For the activation of *redD*, a single colony was then restreaked on MYG plates supplemented with 50 µg/mL apramycin and incubated at 30 °C for three days. For XylE assay, a single colony was inoculated into 25 mL of MYG liquid medium supplemented with 50 µg/mL apramycin in a 125 mL shake-flask (3 mm glass beads added to improve liquid mixing and aeration). Cells were harvested after 72 h cultivation at 30 °C, 250 rpm for XylE assay. Oxytetracycline was added to a final concentration of 6 µM after 6 h growth for inducible repression.

6.4.3 XylE Assay

Aliquots of 10 mL of the cell culture were collected and centrifuged at 4000 rpm for 10 min to remove the supernatant. Five milliliters of sample buffer (100 mM phosphate buffer, pH7.5; 20 mM Na-EDTA, pH 8.0; 10% v/v acetone) was added to resuspend the cell pellets followed by sonication on ice for 1 min (60% amplitude, 15 s on / 30 s off). 0.1%Triton X-100 solution was added into the samples, which were then incubated on ice for 15 min. Then, the samples were centrifuged for 5 min (40,000*g*, 4°C), and the cell lysates were transferred to fresh tubes. 900 µL assay buffer (10 mM phosphate buffer, pH 7.5; 0.2 mM catechol 2,3-dioxygenase) was added to 100µL of the cell lysate, and the change in absorbance at 375 nm (A375) was followed. The concentration of the total proteins was determined by the Bradford method (19). The slope of the linear part of the spectrophotometric output was used to calculate the specific activity of XylE as follows: $mU_{\text{catechol dioxygenase}}$ [nmol min^{-1}] = 30.03×ΔA375/t [min].

6.5 Figures

Figure 6.1. CRISPRi effectively represses the expression of *xylE* in *S. lividans* 66 and *S. albus* J1074. (A) Location of the sgRNAs for CRISPRi. (B) The normalized activity of XylE in the crude extracts of *S. lividans* 66 and *S. albus* J1074.

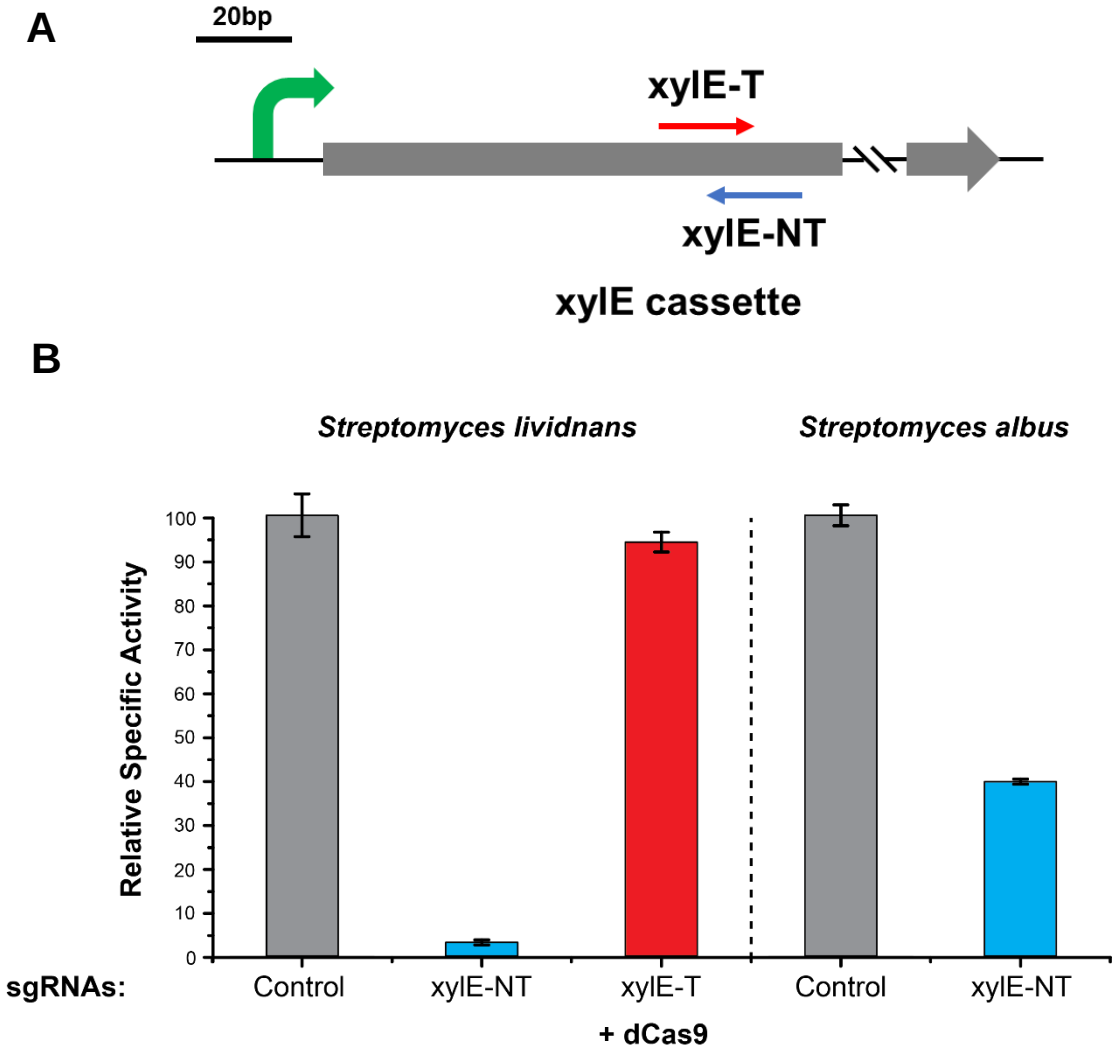


Figure 6.2. CRISPRa improves the productivity of undecylprodigiosin in *S. lividans* 66. (A) Location of the sgRNAs for CRISPRa. (B) *S. lividans* 66 transformants on ISP2 plates. The strain only expressing the fusion protein without sgRNA was used as a negative control.

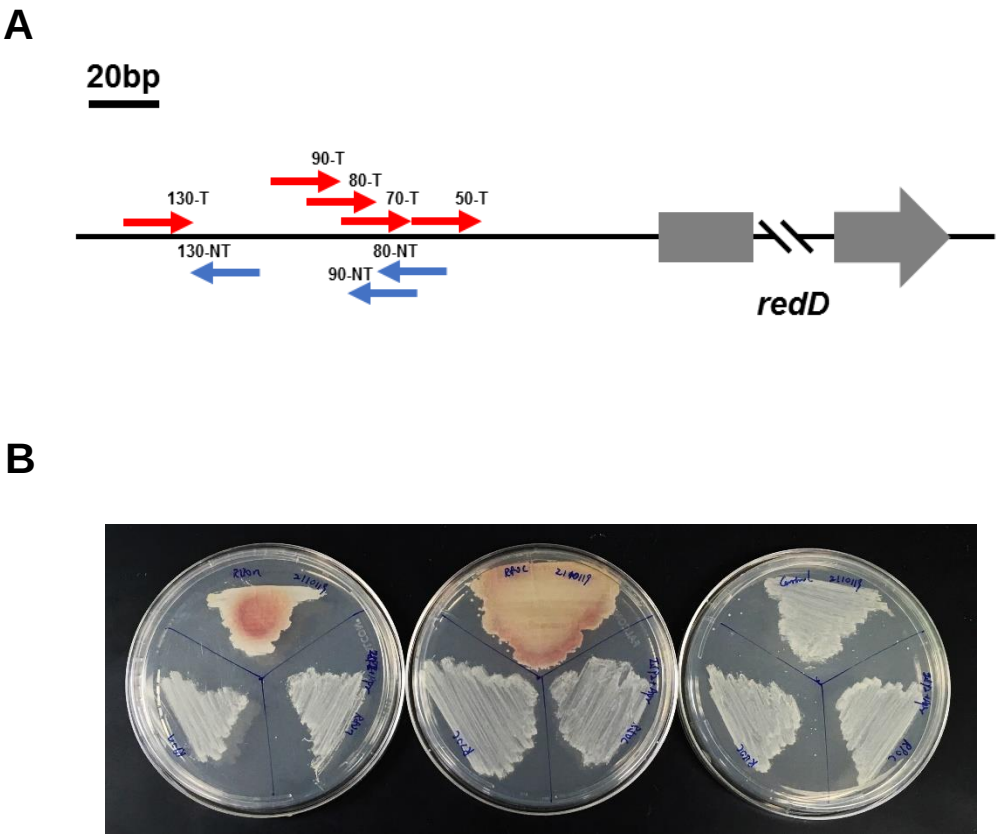
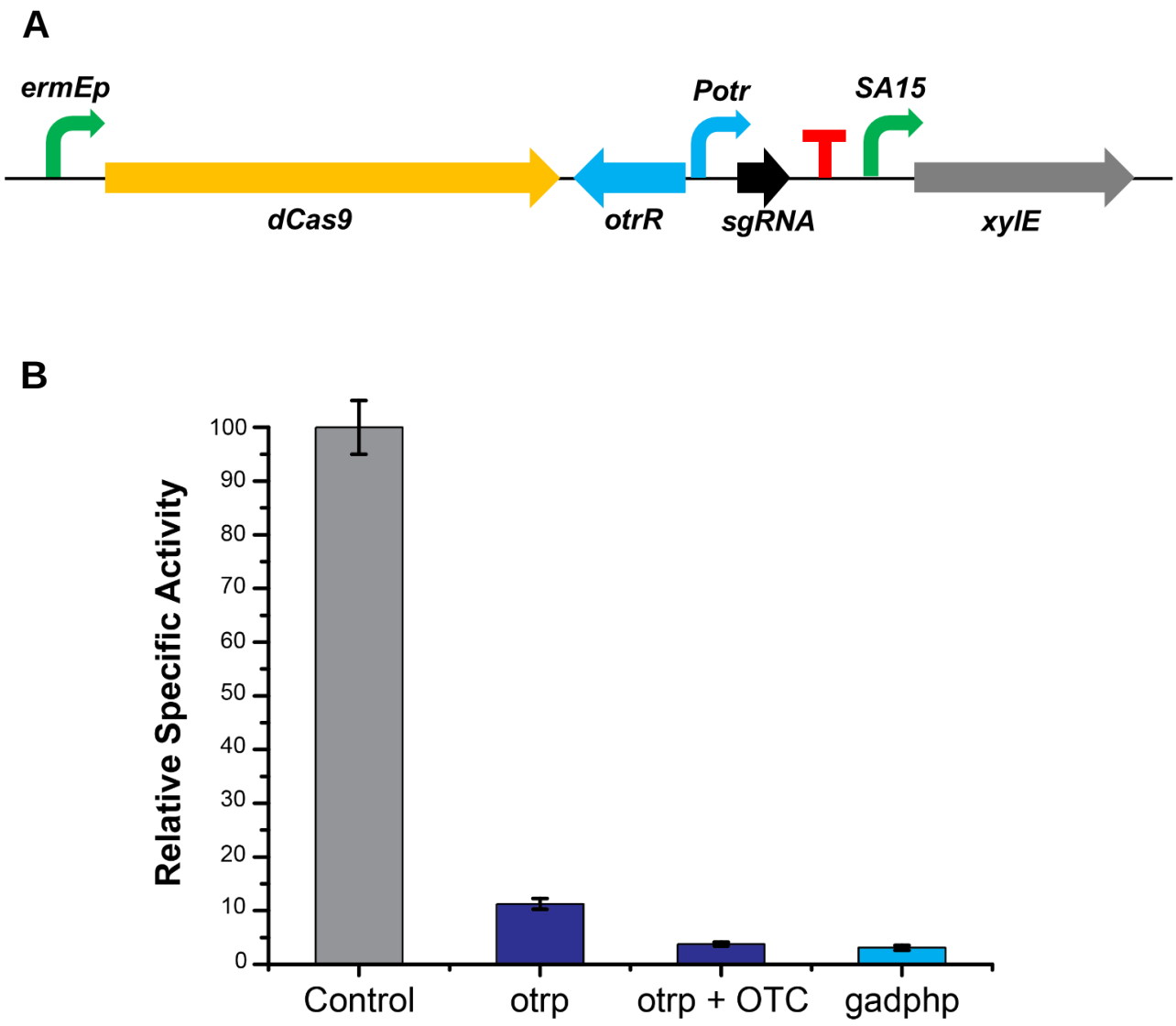


Figure 6.3. Titratable CRISPRi controls gene expression in *S. lividans* 66. (A) Design of the titratable CRISPRi. (B) The normalized activity of XylE in the crude extracts of *S. lividans* 66.



6.6 References

1. Liu R, Deng ZX, & Liu TG (2018) Streptomyces species: ideal chassis for natural product discovery and overproduction. *Metab Eng* 50:74-84.
2. Bentley SD, *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417(6885):141-147.
3. Bilyk O & Luzhetskyy A (2016) Metabolic engineering of natural product biosynthesis in actinobacteria. *Curr Opin Biotech* 42:98-107.
4. Hsu PD, Lander ES, & Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157(6):1262-1278.
5. Doudna JA & Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* 346(6213):1258096.
6. Qi LS, *et al.* (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152(5):1173-1183.
7. Bikard D, *et al.* (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* 41(15):7429-7437.
8. Gilbert LA, *et al.* (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159(3):647-661.
9. Chavez A, *et al.* (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* 12(4):326.
10. Konermann S, *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517(7536):583-U332.
11. Cobb RE, Wang YJ, & Zhao HM (2015) High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS Synth Biol* 4(6):723-728.
12. Huang H, Zheng GS, Jiang WH, Hu HF, & Lu YH (2015) One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta Bioch Bioph Sin* 47(4):231-243.
13. Tong YJ, Charusanti P, Zhang LX, Weber T, & Lee SY (2015) CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth Biol* 4(9):1020-1029.
14. Zeng H, *et al.* (2015) Highly efficient editing of the actinorhodin polyketide chain length factor gene in Streptomyces coelicolor M145 using CRISPR/Cas9-CodA(sm) combined system. *Appl Microbiol Biot* 99(24):10575-10585.
15. Zhao YW, *et al.* (2018) CRISPR/dCas9-mediated multiplex gene repression in Streptomyces. *Biotechnol J* 13(9):1800121.
16. Gordon GC, *et al.* (2016) CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *Metab Eng* 38:170-179.
17. Wang WS, *et al.* (2016) Development of a synthetic oxytetracycline-inducible expression system for Streptomycetes using de novo characterized genetic parts. *ACS Synth Biol* 5(7):765-773.
18. Gomez-Escribano JP & Bibb MJ (2014) Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. *J Ind Microbiol Biot* 41(2):425-431.

19. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248-254.